EPS 1/RM/34 – December 1999
Method Development and Application Section
Environmental Technology Centre
Environment Canada

ENVIRONMENTAL
PROTECTION
SERIES

Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology

Environment Canada
Environnement Canada

Canada

ENVIRONMENTAL PROTECTION SERIES

Sample Number:

EPS 3 / HA / 1

1 — Report number with the qualifier EPS 3/HA
HA — Subject Area Code
3 — Report Category
EPS — Environmental Protection Series

Categories

1	Regulations/Guidelines/Codes of Practice
2	Problem Assessments and Control Options
3	Research and Technology Development
4	Literature Reviews
5	Surveys
6	Social, Economic and Environmental Impact Assessments
7	Surveillance
8	Policy Proposals and Statements
9	Manuals

Subject Areas

AG	Agriculture
AN	Anaerobic Technology
AP	Airborne Pollutants
AT	Aquatic Toxicity
CC	Commercial Chemicals
CE	Consumers and the Environment
CI	Chemical Industries
FA	Federal Activities
FP	Food Processing
HA	Hazardous Wastes
IC	Inorganic Chemicals
MA	Marine Pollutants
MM	Mining and Ore Processing
NR	Northern and Rural Regions
PF	Paper and Fibres
PG	Power Generation
PN	Petroleum and Natural Gas
RA	Refrigeration and Air Conditioning
RM	Reference Methods
SF	Surface Finishing
SP	Oil and Chemical Spills
SRM	Standard Reference Methods
TS	Transportation Systems
TX	Textiles
UP	Urban Pollution
WP	Wood Protection/Preservation

New subject areas and codes are introduced as they become necessary. A list of EPS reports may be obtained from Environmental Protection Publications, Environmental Protection Service, Environment Canada, Ottawa, Ontario, K1A 0H3, Canada

102-04813/

Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology

Method Development and Application Section
Environmental Technology Centre
Environment Canada
Ottawa, Ontario

Report EPS 1/RM/34
December 1999

Canadian Cataloguing in Publication Data

Main entry under title :

Guidance document on application and interpretation of
single-species tests in environmental toxicology

(Report ; EPS 1/RM/34)
Issued also in French under title: Guide des essais
écotoxicologiques employant une seule espèce et de
l'interprétation de leurs résultats.
Includes an abstract in French.
Includes bibliographical references.
ISBN 0-660-16907-X
Cat. no. En49-24/1-34E

1. Environmental toxicology -- Methodology.
2. Toxicity testing -- Methodology.
3. Toxicology -- Methodology.
4. Environmental toxicology -- Canada.
I. Environmental Technology Centre (Canada). Method
Development and Applications Section.
II. Canada. Environment Canada.
III. Series: Information report (Canada. Environment Canada) ;
EPS 1/RM/34.

RA1226.G84 2000 615.9'02'0287 C00-9800100-6

© Minister of Public Works and Government Services 1999
ISBN 0-660-16907-X
Cat. no. En49-24/1-34E

Readers' Comments

Comments regarding the content of this report should be addressed to:

Richard P. Scroggins
Method Development and Application Section
Environmental Technology Centre
Environment Canada
3439 River Road South
Gloucester, Ontario
K1A 0H3

Cette publication est aussi disponible en français. Pour l'obtenir, s'adresser à:

Publications de la Protection de l'environnement
Environnement Canada
Ottawa (Ontario)
K1A 0H3

Review Notice

This report has been reviewed by the staff of the Environmental Technology Advancement Directorate, Environment Canada, and approved for publication. Mention of trade names or commercial products does not constitute recommendation or endorsement by Environment Canada for use.

Abstract

This intent of this document is to provide guidance on using and interpreting single-species toxicity tests for protecting the environment. It focuses on the toxicity tests published by Environment Canada, and supplements the other documents of that department, which deal with sampling, testing, and analyses. Aquatic tests receive most emphasis because they are in more frequent use, but tests with sediments and soil also receive coverage. Both lethal and sublethal tests are included.

The document reviews many topics in environmental toxicology, and might be useful to a person who wished to become familiar with the field. References are given to useful literature on each topic.

Basic concepts of environmental toxicology are explained, with some background on the methods available. Current uses of quality control procedures are outlined, as are methods of dealing with anomalous or troublesome results.

The document attempts to show how toxicity results complement other aspects of environmental work, including broadly-based field studies. Topics include the roles of different survey tools that are available, methods for attempting to extrapolate laboratory toxicity to the field, mass loading of toxicity, and ancillary conditions that modify test results. Some coverage is given of associated topics for specialized investigation, including prediction of toxicity from chemical structure of molecules, and identification of toxic components in a mixed effluent.

Brief reviews are provided of some relevant programs of Environment Canada, including industrial effluent regulation, Environmental Effects Monitoring, oceanic disposal, environmental impact assessment, assessing and improving contaminated sites, the methods for developing guidelines of environmental quality, regulation of new substances for the market, assessment of priority substances under the Canadian Environmental Protection Act (CEPA), and environmental rating of consumer products.

Brief introductions are provided to special topics that show promise for use in environmental toxicology, but are not yet in standard use. These include genetic damage by pollutants, micro-biological testing, multi-species tests, accumulation of toxicants within organisms, biochemical and physiological tests, and automated biological tests for monitoring. Some specific examples of using toxicity tests are given in appendices.

Résumé

Le présent document a pour but d'aider à utiliser et à interpréter les résultats d'essais toxicologiques portant sur une seule espèce afin de protéger l'environnement. Insistant sur les essais publiés par Environnement Canada, il complète les autres documents de ce ministère qui portent sur l'échantillonnage, la conduite des essais et les analyses. Il apporte sa plus grande attention aux essais en milieu aqueux, plus fréquemment utilisés, sans négliger les essais avec les sédiments et le sol. Le document englobe les essais sur les effets létaux et sublétaux.

Abordant de nombreux aspects de l'écotoxicologie, le document pourrait être utile à ceux qui voudraient se familiariser avec le domaine. Une bibliographie utile se rapporte à chaque sujet.

L'explication des rudiments de l'écotoxicologie est éclairée d'un aperçu des méthodes disponibles, tandis que sont esquissées les marches à suivre pour maîtriser la qualité et les méthodes à opposer aux résultats anormaux ou troublants.

Le document tente de montrer comment les résultats toxicologiques complètent d'autres aspects des travaux sur l'environnement, y compris les études non spécialisées sur le terrain. Il traite de sujets tels que le rôle des différentes méthodes de relevé, les méthodes pour extrapoler du laboratoire au terrain les résultats des essais toxicologiques, l'expression de la toxicité en unités de charge massique et les conditions accessoires modifiant les résultats des essais. Il aborde des sujets connexes d'enquêtes spécialisées, notamment la prédiction de la toxicité à partir de la structure des molécules et l'identification des constituants toxiques d'un effluent complexe.

On y effectue un bref survol de certains programmes pertinents d'Environnement Canada, notamment les règlements sur les effluents industriels, la surveillance des effets sur l'environnement, l'immersion en mer, l'évaluation des répercussions environnementales, l'évaluation et l'assainissement des lieux contaminés, les méthodes d'élaboration des lignes directrices sur la qualité de l'environnement, la réglementation des nouvelles substances destinées à la commercialisation, l'évaluation des substances d'intérêt prioritaire sous le régime de la Loi canadienne sur la protection de l'environnement *(LCPE) et l'évaluation environnementale des produits de consommation.*

On y trouve un bref aperçu de sujets particuliers, prometteurs pour l'écotoxicologie, mais encore méconnus de l'usage courant, notamment : effets génétiques des polluants, essais microbiologiques, essais portant sur plusieurs espèces, bioaccumulation des toxiques, essais biochimiques et physiologiques, essais biologiques automatisés pour les besoins de la surveillance. Les annexes renferment des exemples précis de l'emploi des essais toxicologiques.

Table of Contents

Abstract v
Résumé vi
List of Tables xi
List of Figures xii
List of Abbreviations and Chemical Formulae xiii
Glossary xiv
Acknowledgements xxii

Section 1
Introduction 1
1.1 Purpose and Scope of this Guidance Document 1
1.2 Toxicity Tests Within a Triad of Techniques 2
1.3 Single-species Tests and their Validity 4
1.4 Benefits and Limitations of Toxicological Approaches 4
1.5 Support for Single-species Tests 7

Section 2
Basic Concepts of Environmental Toxicology 9
2.1 Terminology 9
2.2 Common Features and Design of Tests 10
2.3 Organisms Used in Tests 12
2.3.1 Desirable Features of Test Species 12
2.3.2 Sensitivity of Test Organisms 13
2.3.3 Choice of Species for Local Relevance 14
2.4 Materials Tested 15
2.5 Estimation of Endpoints from Concentration and Effect 15
2.5.1 Quantal Tests 15
2.5.2 Estimating Inhibiting Concentration in Quantitative Tests 17
2.5.3 Estimating NOEC in Quantitative Tests 18
2.6 Precision of Tests and Variation Between Laboratories 20

Section 3
Test Methods and Quality Assurance 22
3.1 Quality Assurance in Toxicity Testing 22
3.1.1 Related International Standards of Laboratory Practice 23
3.2 Test Methods 24
3.2.1 The Standard Pattern of Environment Canada's Test Procedures 24
3.2.2 Reference Toxicants 25
3.2.3 Evaluating Other Methods 27
3.3 Test Reports 28
3.3.1 The Environment Canada Pattern for Good Test Reports 28
3.3.2 Using a Report to Evaluate the Reliability of a Test 29

Section 4
Interpretation of Toxicological Data 30
4.1 Field Validation 30
4.1.1 Validations in Natural Waterbodies 30
4.1.2 Validations in Mesocosms 32
4.1.3 Literature Reviews 33
4.1.4 Design of a Validation Study 34
4.2 Extrapolation of Toxicity Results 35
4.2.1 Application Factors 36

4.2.2 Acute-to-chronic Ratios 37
4.2.3 Factors Based on Wider Data Bases 38
4.2.4 Statistical Extrapolations 38
4.2.5 Comparison of Extrapolation Methods 40
4.3 Toxic Units and Toxic Emissions 40
4.3.1 The Toxic Units Concept 41
4.3.2 Toxic Units and Effects of Mixtures 42
4.3.3 Mass Loading of Toxicity 43
4.4 The "Weight-of-evidence" Approach 44
4.4.1 Sediment Quality Triad 45
4.5 Toxicity Indices 47
4.6 Structure-activity Relationships 49
4.6.1 Types of Structure-activity Relationships 49
4.6.2 Narcotic and Non-narcotic Models 50
4.6.3 Use of Structure-activity Relationships 51
4.7 Toxicity Identification and Reduction 52
4.8 Characterization of Effluents in Sweden 56
4.9 Troublesome Sets of Results 57
4.9.1 Potential Anomalies in Data-sets 57
4.9.2 Dealing with Problems in Test Results 60
4.10 Toxicity of Effluents 61
4.10.1 Monitoring Effluents 62
4.10.2 Predicting Water Quality and Regulating Toxic Effects 62
4.10.3 Recovery of Damaged Aquatic Ecosystems 66
4.11 Toxicity of Sediment 67
4.11.1 Tests with Whole Sediment 68
4.11.2 Tests with Porewater or Extract 71
4.11.3 Normalization for Total Organic Carbon 73
4.11.4 Normalization for Acid-volatile Sulphide 74
4.12 Toxicity of Soils 75
4.12.1 Tests with Whole Soil 76
4.12.2 Selecting Reference Soils 77
4.12.3 Tests on Soil Extracts 78
4.13 Toxicity of Substances 79
4.13.1 Testing Substances with Troublesome Properties 79
4.14 Toxicity of Surface Waters 80
4.14.1 Laboratory Tests of Receiving Water 81
4.14.2 *In-situ* Tests 81
4.14.3 Sublethality Beyond a Mixing Zone 82
4.15 Conditions that Modify Toxicity 82
4.15.1 Abiotic Modifiers 83
4.15.2 Biotic Modifiers 89

Section 5
Applications of Toxicological Data 92
5.1 Control of Effluents 92
5.2 Sublethal Toxicity Data in Environmental Effects Monitoring 93
5.2.1 Validation of Test Results 94
5.2.2 Tabulation of Sublethal Toxicity Test Endpoints 96
5.2.3 Monitoring Temporal Changes in Effluent Quality 96
5.2.4 Estimating Potential Effects in the Receiving Environment 98
5.2.5 Understanding Effects in Multiple Discharge Situations 100

5.2.6 Integrating Toxicity Results into Environmental Monitoring 101
5.2.7 Mixing Zones 102
5.3 Assessing Sediment Toxicity for Dredging and Disposal at Sea 103
5.3.1 Canadian Approach 103
5.3.2 Reference Sediments 104
5.3.3 Assessing Waste Sediments for Toxicity 107
5.3.4 Monitoring Disposal Sites for Toxicity 109
5.4 Environmental Assessments 109
5.4.1 The General Process 110
5.4.2 Toxicity Tests 112
5.4.3 Public Presentation of Toxicity Information 112
5.5 Assessment and Remediation of Contaminated Sites 113
5.5.1 Ecological Risk Assessment 114
5.5.2 Toxicity Tests for Assessing Contaminated Sites 115
5.5.3 Toxicity Testing in Remediation of Sites 118
5.5.4 Biological Guidelines for Remediation of Contaminated Sediment 118
5.5.5 Allowing for Uncertainty 120
5.6 Development of Environmental Quality Guidelines 120
5.6.1 Background 121
5.6.2 Water Quality Guidelines 122
5.6.3 Sediment Quality Guidelines 125
5.6.4 Soil Quality Guidelines 126
5.6.5 Development of Environmental Quality Objectives 127
5.7 Assessment and Regulation of New Substances 129
5.7.1 Requirements for Data 130
5.7.2 Interpreting Data for Environmental Toxicity 130
5.7.3 Establishing the Reliability of the Toxicity Profile 130
5.7.4 Determining the Need for Further Studies 134
5.7.5 Extrapolation to Concentrations of Concern 135
5.8 Assessment of Priority Substances under CEPA 135
5.8.1 Background 137
5.8.2 Environmental Assessment 137
5.8.3 General Recommendations for Characterization of Effects 138
5.8.4 Types of Data Used in Characterization of Effects 138
5.8.5 Deriving Critical Toxicity Values 139
5.8.6 Types of Biota and Environmental Compartments of Concern 140
5.9 Evaluation of Consumer Products 141
5.9.1 Toxicity Testing in Product Evaluation 142

Section 6
Emerging Issues 144
6.1 Genotoxicity 144
6.1.1 Tests Measuring Damage to DNA/RNA 145
6.1.2 Teratogenicity 146
6.1.3 Surveys of Tumours 146
6.2 Microscale Toxicity Tests 147
6.2.1 Examples of Tests Available 148
6.3 Tests with Microcosms and Mesocosms 149
6.3.1 Tests with Microcosms 150
6.3.2 Tests with Mesocosms 151
6.4 Body Burdens 152
6.5 Mixed-function Oxidases 154

6.6 Endocrine Disrupters 156
6.7 Continuous On-line Biomonitoring 157

References 159

Appendix A
Members of the Project Steering Committee, Development of a Guidance Manual on Environmental Toxicological Data Interpretation and Application 192

Appendix B
The PEEP and CHIMIOTOX Tests Used in the Effluent Testing Program of the St. Lawrence Action Plan 193

Appendix C
An Example of Toxicity Testing as a Predictive Tool in Environmental Planning 196

Appendix D
Example Calculations for Predicting Potential Zones of Sublethal Effect in an Effluent Plume 200

List of Tables

1 Benefits and Limitations of Toxicity Testing 6

2 Summary of Reference Toxicants used in Environment Canada Test Methods 26

3 Comparison of Extrapolations by Different Methods, to Concern Levels for Eight Chemicals 41

4 Interpretation of Effects in the Sediment Quality Triad 48

5 Selected Examples of Quantitative Structure-activity Relationships (QSARs) for Some Classes of Chemicals 53

6 Percentage of Ammonia Un-ionized in Distilled Water 85

7 Acceptable Sublethal Toxicity Tests for Environmental Effects Monitoring 95

8 Minimum Reporting under EEM, for Sublethal Tests with *Ceriodaphnia dubia* 97

9 Toxicity Tests Recommended for the National Contaminated Sites Remediation Program 117

10 The Minimum Set of Toxicity Tests Required for Deriving a Canadian Water Quality Guideline for Aquatic Life 124

11 Exposure Pathways and Receptors Considered in Deriving Soil Quality Criteria for the Four Categories of Land Use 128

12 Physical, Chemical, and Toxicological Information Required for Notification of New Substances Under the Canadian Environmental Protection Act 131

List of Figures

1 Results of a Lethal Test Plotted on Logarithmic-probability Paper 16

2 Warning Chart for Tests with a Reference Toxicant 27

3 Agreement of Predictions from Toxicity Tests on Effluents, with Observed Impact on Receiving-water Communities 33

4 Use of Tri-axial Diagrams to Illustrate Results from the Sediment Quality Triad Applied to Metal-contaminated Sediments 46

5 A Quantitative Structure-activity Relationship (QSAR) for Aliphatic Alcohols 52

6 Sequences of Manipulations of a Sample in Phase I of a Toxicity Identification Evaluation 55

7 Unusual Concentration-effect Curves 58

8 Empirical Relation Between Copper Toxicity and Hardness of Fresh Water 87

9 A Framework for Assessing Wastes for Disposal at Sea 105

10 Canadian Approach for Monitoring Marine Sites used for Disposal of Wastes 107

11 General Model of an Environmental Assessment 111

12 The Protocol of the Canadian Council of Ministers of the Environment for Deriving Water Quality Guidelines for the Protection of Aquatic Organisms . . . 123

13 Sequential Risk-assessment Procedure and Predicted Margin of Safety 132

14 Toxicity Curves for Two Hypothetical Toxicants 134

15 Hypothetical Situations in which Predicted Margins of Safety are Modified by Conditions Existing in the Field 136

List of Abbreviations and Chemical Formulae

AETE	Aquatic Effects Technology Evaluation Program
AHH	aryl hydrocarbon hydroxylase
APHA	American Public Health Association
ASTM	American Society for Testing and Materials
AVS	acid-volatile sulphide
BOD	biochemical oxygen demand
CAEAL	Canadian Association for Environmental Analytical Laboratories Inc.
$CaCO_3$	calcium carbonate
CCREM	Canadian Council of Resource and Environment Ministers
CCME	Canadian Council of Ministers of Environment
CEC	cation exchange capacity
CEPA	Canadian Environmental Protection Act
COD	chemical oxygen demand
CV	coefficient of variation
d	day(s)
DFO	Department of Fisheries and Oceans
DNOC	4,6-dinitro-orthocresol
DO	dissolved oxygen
DOC	dissolved organic carbon
DOM	dissolved organic matter
EA	environmental assessment
ECOD	7-ethoxycoumarin *o*-de-ethylase
ECP	Environmental Choice Program
EC50	median effective concentration
EDTA	ethylene diamine tetra-acetic acid
EEC	European Economic Community
EEM	Environmental Effects Monitoring
Eh	oxidation–reduction (redox) potential
EPS	Environmental Protection Service
ERA	Ecological Risk Assessment
EROD	ethoxyresorufin *o*-de-ethylase
FeS	iron monosulphide
g	gram(s)
GLP	Good Laboratory Practices
h	hour(s)
HPLC	high-pressure liquid chromatography
IC	inhibiting concentration
IJC	International Joint Commission
IR	infrared
ISO	International Organization for Standardization
kg	kilogram(s)
K_{oc}	organic carbon partition coefficient
K_{ow}	octanol–water partition coefficient
L	litre(s)
LC50	median lethal concentration
LOEC	lowest-observed-effect concentration
LTU	lethal toxic units
MFO	mixed-function oxidase(s)
mg	milligram(s)
MISA	Municipal Industrial Strategy for Abatement [Ontario]
mol/L	moles per litre
mw	molecular weight
NCSRP	National Contaminated Sites Remediation Program
NOEC	no-observed-effect concentration
OECD	Organization for Economic Co-operation and Development
PAH	polycyclic aromatic hydrocarbon(s)
PCB	polychlorinated biphenyl(s)
PCPA	Pest Control Products Act
PEEP	Potential Ecological Effects Probe
pH	acid/base reaction of an aqueous solution
p-NT	para-nitrotoluene
QA/QC	Quality Assurance/Quality Control
QSAR	quantitative structure-activity relationship(s)
r	correlation coefficiant
s	second(s)
SAR	structure-activity relationship(s)
SD	standard deviation
SLTU	sublethal toxic units
t	tonne
TDS	total dissolved solids
TEF	toxicity emission factor
TIE	toxicity identification evaluation
TOC	total organic carbon
TOEC	threshold-observed-effect concentration
TRE	toxicity reduction evaluation
TU	toxic unit(s)
USEPA	United States Environmental Protection Agency
UV	ultraviolet
2,4-D	2,4-dichlorophenoxyacetic acid
2,4-DCA	2,4-dichloroacetic acid
>	greater than
<	less than
≥	greater than or equal to
≤	less than or equal to
%	per cent

Glossary

All definitions are given in the context of the procedures in this report, and might not be appropriate in another context.

Grammatical Terms

Can is used to mean "is (are) able to."
May is used to mean "is (are) allowed to".
Might expresses a possibility that something could exist or happen.
Must is used to express an absolute requirement.
Should is used to state that the specified condition or procedure is recommended and ought to be met if possible.

Technical Terms

Accreditation is the formal recognition of the competence of an environmental analytical laboratory to carry out specified tests. Formal recognition is based on an evaluation of laboratory capability, partly judged by site inspections, and performance.

Acute means within a short period (minutes, hours, or a few days) in relation to the life span of the test organism.

Acute toxicity is an adverse effect (lethal or sublethal) induced in the test organisms within a short period of exposure to a test material, usually a few days.

Battery of tests is a combination of several toxicity tests on different organisms. The term implies diverse types of organisms (e.g., tests with a bacterium, an alga, or a crustacean).

Bioaccumulation is the net accumulation of a substance by an organism as a result of uptake from all environmental sources.

Bioassay is a test (= assay) in which the strength or potency of a substance is measured by the response of living organisms, or living system. In standard pharmacological usage, a bioassay assesses the unknown potency of a given preparation of a drug, compared to the known potency of a standard preparation. *Toxicity test (q.v.)* is a more specific and preferred term for environmental work.

Bioconcentration is the net accumulation of a substance by an organism from the surrounding medium, to a concentration higher than in the medium. For an aquatic organism, bioconcentration would result from direct uptake from ambient aqueous solution.

Biomagnification is the increase in tissue concentration of poorly depurated substances by dietary accumulation through two or more trophic levels. (An increase through only one trophic exchange, i.e., from one set of organisms to consumers at the next trophic level, has been termed by some agencies elsewhere as *secondary poisoning*, but the words would be ill-chosen for cases showing an increase in concentration without indication of actual poisoning.)

Chemical is, in this report, any element, compound, formulation, or mixture of compounds that might be mixed with, deposited in, or found in association with soil, sediment, or water.

Chronic means occurring during a relatively long period of exposure, usually a significant proportion of the life span of the organism, such as 10% or more.

Chronic toxicity implies long-term effects that are related to changes in metabolism, growth, reproduction, or ability to survive.

Compliance means in accordance with government regulations or requirements for issuing a permit.

Contaminant is a foreign substance that is present in a natural system, or present at increased concentration, because of some direct or indirect human activity. The term is usually applied to substances that are present in low concentrations, which have not been demonstrated to cause adverse biological effects. Various groups use the term in particular senses, and it has specific meanings under some national and international definitions or regulations.

Control is a treatment in an investigation that duplicates all the factors that might affect results, except the specific condition being studied. In toxicity tests, the control must duplicate all conditions in the exposure treatment(s) but must contain no test material. The control is used as a check for toxicity due to basic conditions such as quality of dilution water or health and handling of organisms. *Control* is synonymous with *negative control*. See also *positive control*.

Control sediment is clean sediment, taken from the site where the test organisms were collected, or formulated to meet a standard composition. For cultured organisms, the control is a sample of sediment identical to that used for the culture. This sediment must contain no test material. Environment Canada test methods require that the effect in control sediment must not exceed some specified pre-determined rate.

Criteria have been defined in Canada as scientific data evaluated to derive the recommended limits for water uses (as applied to water quality by CCREM, 1987). The singular is *criterion*. See also *quality guideline*. (Readers should be aware that in literature from some other places such as the USA, and in common usage by professionals, criterion is used in the sense of *guideline* as defined below. For example, Rand (1995) defines water quality criterion as "an estimate, based on scientific judgments, of the concentration of a chemical or other constituent in water which, if not exceeded, will protect an organism, a ... community, or a prescribed water use or quality with an adequate degree of safety".)

Critical body burden (CBB) is the minimum concentration of a substance in the tissues of an organism, that is associated with an adverse effect in that organism. The term has particular relevance to Priority Substances (Section 5.8).

Critical toxicity value (CTV) is a quantitative expression of low toxic effect, such as an IC25. The expression is used in Canadian assessments of Priority Substances.

Culture, as a noun, means the stock of organisms raised under defined and controlled conditions to produce healthy test organisms. As a verb, it means to carry out the procedure of raising organisms.

Depuration is the loss of a substance from an organism as a result of any active or passive process.

EC50 is the *median effective concentration*. That is the concentration of material in water (e.g., mg/L) or soil or sediment (e.g., mg/kg) that is estimated to cause some defined toxic effect to 50% of the test organisms. In most instances the EC50 and its 95% confidence limits are statistically derived by analyzing the percentages of organisms affected at various test concentrations, after a fixed period of exposure. The duration of exposure must be specified (e.g., 72 h). The EC50 describes quantal effects, lethal or sublethal, and is not applicable to quantitative effects (see *ICp*).

Ecological risk assessment is the process of identifying and quantifying *risks (q.v.)* to nonhuman organisms and determining the acceptability of those risks.

Ecotoxicology has the same general meaning as *environmental toxicology (q.v.)*.

Effluent is any liquid waste (e.g., industrial, municipal) discharged to the environment.

Elutriate is an aqueous solution obtained after adding water to a solid material (e.g., soil, tailings, drilling mud, dredge spoil), shaking the mixture, then centrifuging it, or filtering it, or decanting the supernatant.

Endpoint means the measurement(s) or value(s) that characterize the results of a test (LC50, EC50, IC25, etc.). It also means the reaction of the organism to show the effect which is intended to mark completion of the test (e.g., death, number of young produced). In environmental assessment, particularly for Canadian Priority Substances, endpoint has another meaning, usually within the term *assessment endpoint*; this signifies the species, kind, or category of organism that is being considered for risk of harmful effects. See also *receptor*.

Environmental toxicology is a branch of *toxicology (q.v.)* with the same general definition. However the focus is on ecosystems, natural communities, and wild living species, without excluding humans as part of the ecosystems.

Flow-through describes tests in which solutions in test vessels are renewed continuously by the constant inflow of a fresh solution, or by a frequent intermittent inflow.

Geometric mean is the mean of repeated measurements, calculated on a logarithmic basis. It has the advantage that extreme values do not influence the mean as is the case for an arithmetic mean. The geometric mean can be calculated as the n^{th} root of the product of the "n" values and it can also be calculated as the antilogarithm of the mean of the logarithms of the "n" values.

Good Laboratory Practices is a set of standards governing the experimental design, collection of data, and conduct of scientific and technical studies in the laboratory. The standards are promulgated by the Organization for Economic Co-operation and Development (OECD).

Graded effects. See *quantitative effects*.

Guideline. See *quality guideline*.

Hazard is the potential for adverse effect(s) that might result from exposure(s) to chemicals at potentially toxic concentrations.

Hazard assessment is an initial part of risk assessment. It defines which biotic components are potentially exposed to toxic concentrations, and how they might be affected.

ICp is the *inhibiting concentration for a (specified) percent effect*. It represents a point estimate of a concentration of test material that causes a designated percent inhibition *(p)* compared to the control, in a quantitative biological measurement such as number of young produced or size attained by individuals at the end of the test.

Interstitial water is the water which is in its undisturbed state within the sediment, occupying the space between sediment particles. If extracted from the sediment, it should no longer be called interstitial water. (Porewater *(q.v.)* is a more general term which signifies interstitial water, but also signifies the same water after it has been extracted from the sediment.)

K_{ow} is the octanol–water partition coefficient of an organic chemical, *i.e.*, the ratio of the solubilities of a chemical in *n*-octanol and water. Higher ratios indicate greater tendency to bioaccumulate. Sometimes the symbol P_{oct} is used.

LC50 is the *median lethal concentration*, i.e., the concentration of material in water or soil or sediment that is estimated to be lethal to 50% of the test organisms. The LC50 and its 95% confidence limits are usually derived by statistical analysis of percent mortalities in several test concentrations, after a fixed period of exposure. The duration of exposure must be specified (e.g., 48-h LC50).

Leachate is water or wastewater that has percolated through soil or solid waste.

Lethal means causing death by direct action. Death is usually defined as the cessation of all visible signs of movement or other activity.

Life-cycle test is one in which the organisms are observed at least from a life stage in one generation to the same life stage in the next generation.

LOEC is the *lowest-observed-effect concentration*. This represents the lowest concentration of a test material for which a statistically significant effect on the test organisms was observed, relative to the control.

LT50 is the *median lethal time*, i.e., the exposure time that is estimated to be lethal to 50% of the test organisms for a given concentration of test material. Successive observations of mortality in each of a series of concentrations can allow an estimate of LT50 for each concentration, and sometimes provides a revealing *toxicity curve (q.v.)*. The usual statistical techniques for LC50 are not valid because of repeated observations on each group of organisms.

MATC signifies the "maximum acceptable toxicant concentration", a term that is used mostly in the USA. It has been defined in various ways but is now generally considered to be synonymous with *TOEC*, the term recommended herein.

Material is the *substance (q.v.)* or substances from which a thing is made. A material would have more or less uniform characteristics. Sediment, soil, or surface water would be materials. Usually, the material would contain several or many substances or chemicals.

Narcosis is a non-specific mode of toxic action, a general disruption of cellular activity that can be produced by many organic chemicals. The term is also used to describe the observed effect on animals, generally a state of stupor or insensibility, and loss of equilibrium or movement.

Negative control (See *control*.)

NOEC is the *no-observed-effect concentration*. This represents the highest concentration of a test material at which no statistically significant effect on the test organisms was observed, relative to the control.

Nonpolar substances are those having molecules that are electrically symmetrical, with the centre of positive and negative charges at the same point in the molecule. Normally this is because the constituent atoms of the molecule share electrons, which lie midway between (or among) the atoms. A nonpolar solvent such as benzene or carbon tetrachloride dissolves similar nonpolar or covalent substances such as sulphur, phosphorus, and waxes, which are insoluble in water. See also *polar*.

Objective. See *quality guideline*.

Pelagic describes an aquatic organism which remains free-swimming or free-floating. It also signifies an open-water environment, or the water between the surface and the floor of the waterbody, generally a sea.

PMTC is the *predicted minimum toxic concentration*, a term developed by Environment Canada for use in monitoring environmental effects of effluents. It is defined as the concentration of an effluent in the receiving water, below which only 5% of samples would be expected to have a deleterious sublethal effect, as estimated at the 95% level of confidence.

P_{oct} is the octanol–water partition coefficient. See K_{ow}.

Polar substances are those having molecules that are not electrically symmetrical, one end of the molecule being positive and the other negative, reflecting the relative distribution of electrons. Water is a polar solvent because its molecules are dipoles, with the constituent atoms (hydrogen and oxygen) asymmetrically arranged so that the centres of positive and negative electrical charges are not located at the same point in the molecule, i.e., one end tends to be positive and the other negative. Accordingly, water is a good solvent for acids, bases, and salts – those substances that are ionic or that produce ions in solution. As a polar solvent, water is almost completely immiscible with liquids such as benzene and carbon tetrachloride which are *nonpolar* (q.v.).

Pollutant is the material or form of energy that causes *pollution (q.v.)*, or is capable of causing pollution if discharged to the environment in sufficient quantities. (See also *contaminant*.)

Pollution is the addition of a material or a form of energy such as heat, to some component of the environment, in such an amount as to cause a detectable change which is deleterious to some organism or to some human use of the environment. Some national and international agencies have formal definitions of pollution, which should be honoured in the appropriate contexts.

Porewater is the general term describing the water which occupies the space between sediment particles. The amount of porewater is expressed as percent of the wet sediment, by weight. Porewater can be extracted from the sediment, and tested for toxicity, if desired. (Interstitial water refers to porewater that is in its undisturbed state within the sediment; if extracted, it should no longer be called interstitial water.)

Positive control is a toxicity test with a reference toxicant, used to assess the sensitivity of the organisms at the time the test material is evaluated and the precision of results obtained by the laboratory for that chemical.

Priority Substance is a material that has been included in the *Priority Substances List* by the federal Ministers of Environment and Health, under a requirement of the Canadian Environmental Protection Act. The material must be assessed to determine whether it is toxic and whether mitigation is required.

Probit is a unit of divergence from the mean of a normal distribution, expressed in terms of the standard deviation of the distribution. Its practical use in estimating an LC50 or EC50 is in straightening the sigmoid curve of the normal distribution, for percent effect as a function of logarithm of concentration. *Probit analysis* is the procedure using maximum likelihood to estimate the parameters of the normal tolerance distribution, and hence the EC50. In modern usage, "probit analysis" is often used to describe estimation by maximum likelihood, with any underlying distribution of tolerance (e.g., normal, logistic).

Protocol is an explicit set of procedures for a test or experiment, formally agreed upon by the parties involved, and described precisely in a written document.

Quality Assurance (QA) is a program within a laboratory, intended to provide precise and accurate results in scientific and technical work. It includes selection of proper procedures, sample collection, selection of limits, evaluation of data, *quality control (q.v.)*, and qualifications and training of personnel.

Quality Control (QC) consists of specific actions within the program of quality assurance. It includes standardization, calibration, replication, control samples, and statistical estimates of limits for the data.

Quality guideline is a scientifically based numerical concentration limit or narrative statement recommended to support and protect designated indigenous organisms or a specified water use. This could be a soil quality guideline, water quality guideline, etc. A *quality objective* has the same definition as guideline except that it applies to a specific site. Some provinces have established lists of water quality objectives, and they reflect "officially desired conditions". A *quality standard* is an objective that is recognized in enforceable environmental control laws of a level of government.

Quantal effects in a toxicity test are those in which each test organism either responds or does not respond. For example, an animal might either produce young or fail to produce young. See also *quantitative effects*.

Quantitative effects in a toxicity test are those in which the measured effect is continuously variable on a numerical scale. Examples would be number of algal cells counted or size attained by an animal. *Graded* means the same thing in this context. See also *quantal*.

Rearing means collecting test organisms from the wild and maintaining them in the laboratory until they are the right size, age, or weight for toxicity testing, or until they produce healthy offspring.

Receiving water is surface water (e.g., in a stream, river or lake) that has received a waste or contaminant, or else is about to receive such a waste (e.g., it is just upstream from a point source of contamination). Further description must be provided to indicate the meaning intended.

Receptor, as used in environmental assessment, and especially in ecological risk assessment, means a component of the ecosystem (e.g., predatory birds) that could be affected by the toxic substances being considered for a given site or situation. *Receptor characterization* attempts to identify those components.

Reference sediment is a field-collected sample of sediment that has properties (e.g., particle size, total organic content) that closely match those of the sample(s) of test sediment except that it is selected so as to be free from the particular source of chemical contamination that is of interest and is being assessed. It is often selected from a site near the survey stations where samples of test sediment are collected, and thus might be subject to other influences from pollution aside from the one(s) being assessed.

Reference soil is a field-collected sample of soil that has properties (e.g., texture, structure, pH, organic content) that are as similar as possible to those of the sample(s) of test soil except that it is free from the source of chemical contamination that is of interest and is being assessed. It is often selected from the general vicinity of the locations of samples of test soil, and thus might be subject to other influences from pollution aside from the one(s) being studied. It is used to describe matrix effects in the test, and may also be used as a control and as a diluent to prepare concentrations of the test soil.

Reference toxicant is a standard chemical used to measure the sensitivity of the test organisms to establish confidence in the toxicity data obtained for a test material. In most instances, a toxicity test with a reference toxicant is performed to assess the sensitivity of the organisms at the time the test material is evaluated, and the precision of results obtained by the laboratory for that chemical.

Remediation is the management of a contaminated site to prevent, minimize, or mitigate damage to human health or the environment. Remediation can include both direct physical actions (e.g., removal, destruction, and containment of toxic substances) and institutional controls (e.g., zoning designations or orders).

Replicate is a single test chamber containing a prescribed number of organisms in either one concentration of test solution or in dilution water as a control. In a toxicity test comprising five test concentrations and a control, and using three replicates, 18 test chambers would be used. For each concentration or control, there would be three test chambers or replicates. A replicate must be an independent test unit, and therefore, any transfer of organisms or test material from one replicate to another would invalidate a statistical analysis based on the replication.

Resuspended sediment is sediment that has been mixed vigorously with water and allowed to settle.

Risk is the probability or likelihood that an adverse effect will occur.

Risk assessment See *ecological risk assessment*.

"Safe" has many possible definitions. Here, when used with quotation marks, it signifies a concentration of toxicant or waste material, *just* low enough that it would cause no sublethal harm to living organisms upon an indefinitely prolonged exposure. Such a "safe" concentration is accordingly somewhat hypothetical and without a defined arithmetic method of estimation, but would be lower than any concentration associated with toxic effects.

Salinity has traditionally referred to measurements of the total mass of dissolved salts in a given mass of solution, described in terms of g/kg or "parts per thousand" (‰). Today it is empirically measured from standard relationships of density or conductivity, and results are stated without units, for example, "salinity is 32" (APHA *et al.*, 1995).

Sediment is a natural particulate material that has been transported to, and deposited at, the bottom of a body of water. The term can also describe a substrate that has been experimentally prepared. The term refers to whole, intact sediment that has had minimal manipulation following collection or formulation. It does not refer to a form or derivative of the sediment such as an elutriate or a resuspended sediment. Except in situations where special emphasis is required, it is not necessary to write *whole sediment*. Material which had settled and become sediment might be re-mixed into the water, in which case it would be called *suspended sediment*.

Sediment elutriate. See *elutriate*.

Serial dilution describes a series of test concentrations in which each lower concentration differs from the preceding one by a constant factor (dilution), such as 100, 50, 25, 12.5%. Such a series can be obtained by successive dilutions of a given wastewater or stock solution.

Soil is the whole, intact material that has had minimal manipulation following collection or formulation. It is not a form or derivative such as an elutriate or a leachate. Except in situations where special emphasis is required, it is not necessary to write *whole soil*.

Soil elutriate. See *elutriate*.

Spiking refers to the addition of a known amount of chemical or other material to a sediment. The material is usually added to a clean or control sediment, but sometimes to a contaminated sediment. After the addition, the sediment is mixed thoroughly. If the added material is a test sediment, Environment Canada documents do not call this spiking, but instead "dilution" or simply "addition".

Standard. See *quality guideline*.

Static describes toxicity tests in which test solutions or water overlying test sediment are/is not renewed during the test.

Static-renewal describes toxicity tests in which test solutions or water overlying test sediment are/is renewed (replaced) periodically during the test, usually at the beginning of each 24-h period or less. Synonymous terms are "renewal", "batch replacement" and "semi-static".

Sublethal means detrimental to the organism, but below the level that directly causes death within the test period.

Substance is a particular kind of material. The word substance has a narrower scope than *material (q.v.)*, and might refer to a particular chemical or element.

Suspended sediment is sediment, or part of it, that has become mixed with water, with the particles continuing to be supported in the water.

TOEC is the *threshold-observed-effect concentration*. Its true value lies somewhere between the NOEC and LOEC, and is estimated as the geometric mean of those two concentrations, for the convenience of having a single endpoint.

Toxic means poisonous. A toxic material can cause adverse effects on living organisms, if present in sufficient amount at the right location. Toxic is an adjective or adverb, and should not be used as a noun. *Toxicant* is a legitimate noun. (See also Section 2.1.)

Toxicant is a toxic material.

Toxic emission factor (TEF) is the *toxic emission rate (q.v.)* for a given effluent, per unit production of goods (e.g., one tonne of paper) at the operation which generates the effluent.

Toxic emission rate (TER) is the toxic strength of an effluent multiplied by the volume discharged. The value of toxic units for the effluent would be multiplied by the discharge in m^3/d.

Toxic unit is an expression of the toxic strength of a substance in solution, or of a wastewater. Strength is expressed as a multiple of (= fraction of) a standard endpoint of toxicity. For example, *lethal toxic units* of a chemical in a solution would be calculated as the actual concentration of the chemical in the water, divided by the LC50 of the chemical. *Sublethal toxic units* would be calculated by using a defined sublethal endpoint (such as IC25) as the denominator. Toxic units of an effluent would be calculated as 100% (i.e., the strength of the effluent) divided by the measured toxic endpoint as a percentage (e.g., an effluent with an LC50 of 10% would have 10 lethal toxic units). Toxic units are conceptually convenient since their numerical value increases as the toxicity increases.

Toxicity is the inherent potential or capacity of a material to cause adverse effects on living organisms.

Toxicity curve is a graph of successive endpoints of a test or tests, plotted against exposure time, both on logarithmic scales (e.g., log LC50 versus log exposure time). The curve can indicate whether a threshold of toxicity was reached during the test, i.e., a time-independent asymptote of concentration, an important item of knowledge for any toxic chemical. A toxicity curve is usually for lethal effects, since for most sublethal tests, definitive observations of effect are available only at the end of the test.

Toxicity identification evaluation (TIE) is a systematic investigation to discover what substance(s) in a mixture causes the mixture to be toxic. Various physicochemical pretreatments (e.g., pH adjustment, filtration, aeration) are followed by tests for toxicity; the results provide clues to the type of toxicant acting.

Toxicity test is a determination of the effect of a material on a group of selected organisms under defined conditions. An environmental toxicity test usually measures either (a) the proportions of organisms affected (*quantal*), or (b) the degree of effect shown (*graded* or *quantitative*) after exposure to a chemical, liquid waste, receiving water, sediment, or soil, at specific concentrations or in full strength.

Toxicity reduction evaluation is similar to *toxicity identification evaluation (q.v.)*, but designed to illustrate which techniques for waste treatment are most successful in reducing toxicity.

Toxicology is a branch of science that studies the toxicity of materials or conditions. There is no limitation on the use of various scientific disciplines, field of laboratory tools, or studies at various levels, whether molecular or ecosystem. Applied toxicology would normally have a goal of defining the limits of safety of chemical or other agents. (See also *environmental toxicology*.)

Warning limit is plus or minus two standard deviations, calculated on a logarithmic basis, from a historic geometric mean of the endpoints from tests with a reference toxicant.

Warning chart is a graph used to follow changes over time, in the endpoints for a reference toxicant. Date of the test is on the horizontal axis and the effect-concentration is plotted on the vertical logarithmic scale (see Figure 2).

Whole soil. See *soil*.

Whole sediment. See *sediment*.

Acknowledgements

This document was written by John B. Sprague (Sprague Associates Ltd., Salt Spring Island, B.C.), Barry R. Taylor (then Golder Associates Ltd., Calgary, Alta., currently Taylor Mazier Associates, St. Andrews, N.S.), Hal R. Hamilton, Gordon G. Macdonald, Stella M. Swanson, and Michael G. Rankin (Golder Associates Ltd., Calgary, Alta. and Burnaby, B.C.), and J. Steven Goudey (HydroQual Laboratories Ltd., Calgary, Alta.). Section 5.8 was written by Kenneth W. Taylor (Commercial Chemicals Evaluation Branch, Environment Canada, Hull, Que.). Parts of Section 5.2 were prepared by Richard P. Scroggins (Environment Canada). This document is based on previous guidance documents, reports, and the published literature in toxicology and ecology, supplemented by suggestions from scientific and technical people in government, industry, and academia from across Canada. Richard P. Scroggins (Method Development and Application Section, Environment Canada, Gloucester, Ont.), was the Scientific Authority for the project, providing general guidance, detailed proofing, and technical assistance throughout the work.

Members of the Project Steering Committee are all actively involved in environmental toxicology and are thanked for continued participation in the development and review of this document. The members are from Environment Canada unless otherwise mentioned, and are as follows. Robert Baker (Environmental Assessment Branch, Hull, Que.), Norman Bermingham (Centre St. Laurent, Montreal, Que.), Philippa M. Cureton (Commercial Chemicals Evaluation Branch, Hull, Que.), Marie-Louise Geadah (National Office of Pollution Prevention, Hull, Que.), Karen Keenleyside (Evaluation and Interpretation Branch, Hull, Que.), Ian MacKenzie (Standards and Approvals Division, Alberta Environmental Protection, Edmonton, Alta.), Jennifer A. Miller (Miller Environmental Sciences Inc., Stroud, Ont.), James M. Osborne (Marine Environment Division, Hull, Que.), W. Roy Parker (Atlantic Region, Environmental Protection Service, Dartmouth, N.S.), Richard P. Scroggins (Method Development and Application Section, Gloucester, Ont.), Rob Stevens (Phys/Chem. Sciences Directorate, Fisheries and Oceans Canada, Ottawa, Ont.), and Roger Sutcliffe (Commercial Chemicals Evaluation Branch, Hull, Que.).

Invaluable assistance is gratefully acknowledged, for the following people who provided comments and suggestions on drafts at one or more stages of development: Donald J. McLeay (McLeay Environmental Ltd., Victoria, B.C.); Peter M. Chapman (EVS Environment Consultants, North Vancouver, B.C.); Glenn F. Atkinson (Science Policy Branch, Environmental Conservation Service, Environment Canada, Hull, Que., and Environmental and Biological Statistics, Ottawa, Ont.); Gord Balch (Trent University, Peterborough, Ont.); Steven Bay (Southern California Coastal Water Research Project, Westminster, Calif.); Art Beckett (Prairie and Northern Region, Environment Canada, Edmonton, Alta.); Mark A. Bonnell (Bonnell Environmental Consulting, Ottawa, Ont.); Robert M. Burgess (U.S. Environmental Protection Agency, Narragansett, R.I.); G. Allen Burton (Wright State University, Dayton, Ohio); Andrée Chevrier (Marine Environment Division, Environment Canada, Hull, Que.); Gordon R. Craig and Patricia L. Orr (Beak Consultants Ltd., Brampton, Ont.); Kenneth G. Doe (Toxicology Laboratory, Atlantic Region, Environment Canada, Moncton, N.B.); Joseph Greene (Dept of Civil Engineering, Oregon State Univ., Corvallis, Ore.); Isabelle Guay and Marc Sinotte (Service d'évaluation des rejets toxiques, Min. de l'Environnement, Québec, Que.); Mick Hamer (Zeneca Agrochemicals, Bracknell, Berkshire, U.K.); Susan V. Humphrey (Ontario Region, Environment Canada, Downsview, Ont.); Tibor G. Kovacs and Pierre Martel (Pulp and Paper Research Inst. of Canada, Pointe Claire, Que.); Peter F. Landrum (National Oceanic and Atmospheric Administration, Ann Arbor, Mich.); James M. Lazorchak (Environmental Monitoring Systems Laboratory, Cincinnati, Ohio); Deanna L. Lee (British Columbia Ministry of Environment, Lands and Parks, Surrey, B.C.); Karsten Liber (then Lake Superior Research Inst., Univ. of Wisconsin, Superior, Wis., now Toxicology Centre, Univ. Saskatchewan, Saskatoon, Sask.); Donald Mackay (then Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ont., now Trent Univ., Peterborough, Ont.); Louis Martel and Claude Thellen (Écotoxicologie et Développement, Ministère de l'Environnement, Sainte-Foy, Que.); Stanley Megraw (Domtar Research Centre, Senneville, Que.); Cathy A. McPherson (EVS Consultants, North Vancouver, B.C.); Linda Meyers-Schöne (International Technology Corporation, Albuquerque, N.M.); Sylvain Ouellet (Guidelines Division, Environment Canada, Hull, Que.); Lynne Patenaude (Environmental Choice Program, Terra Choice, Ottawa, Ont.; Linda Porebski (Marine Environment Division, Environment Canada, Hull, Que.); Robert Prairie (Centre de technologie Noranda, Pointe-Claire, Qué.); Anne-Marie Prud'homme (Vice-présidence Environnement et Collectivités, Hydro Québec, Montréal, Qué.); John Retallack (NovaCor Chemicals, Red Deer, Alta.); Phillippe Riebel (Beak Consultants Ltd., Dorval, Que.); Robert Roy (B.A.R. Environmental Inc., Guelph, Ont.); Adolfo E. Silva (Petro-Canada, Calgary, Alta.); Brian Steinback (Alberta Newsprint, Whitecourt, Alta.); Maria Tarkpea (Inst. Applied Environmental Research, Universitet Stockholms, Nyköping, Sweden); K.-L. Tay (Environment Canada, Dartmouth, N.S.); Ron Watts (Pacific Environmental Science Centre, Environment Canada, North Vancouver, B.C.); Stewart G. Yee (Pacific Environmental Science Centre, Environment Canada, North Vancouver, B.C.).; and Rick H. Young (MacMillan Bloedel Ltd., Vancouver, B.C.)

Introduction

Key Guidance

- *This document provides guidance on interpreting whole-organism toxicity tests. It focuses on test methods published by Environment Canada.*
- *Interpretation is strengthened by also considering the results of chemical analyses and biological surveys in the field. Toxicity tests and chemical measurements can be used for regulatory or monitoring purposes, or they can be used to predict effects in nature. Biological surveys are never predictive, but provide the final audit of whether there have been effects.*
- *Toxicity tests are the **only** way to measure toxicity. Their chief disadvantage is that they cannot, by themselves, identify the substance causing toxicity.*
- *Canada's pioneering use of toxicity tests to regulate industrial discharges has had noticeable benefits; such tests are now widely supported by international agencies.*

1.1 Purpose and Scope of this Guidance Document

Toxicity tests are becoming increasingly important in environmental protection, and Environment Canada (EC) now provides standard procedures, quality assurance, and guidance on use of tests. Guidance on how to interpret results of laboratory toxicity tests, and in a general way, how to design tests and incorporate them into programs for environmental protection are presented herein.

The standard methods of Environment Canada are emphasized (EC, 1990a–f; 1992a–f; 1997a,b; 1998; 1999), although guidance is pertinent to toxicity tests in general. Similarly, there is specific guidance for national programs of Environment Canada, but also general guidance on the ecological implications of results from toxicity tests. Guidance is intended to be useful to industries for their own assessments of toxicological data, and also to envision how such data fit into government programs. Some understanding of basic principles in toxicity testing is assumed, but this guidance document should be of practical benefit to users of varying experience.

Primarily single-species tests are dealt with herein, measuring toxicity of contaminated soil, sediments, water and wastewater, as well as individual substances contained in those media. This document supplements reports on standard test methods, and guidance documents on sampling, reference toxicants, and statistical analysis (EC, 1990d; 1994c; 1995a in prep., a). The use of toxicity tests in broader programs of environmental assessment is described in general, with some explanation of those broader programs (field validation of laboratory results, environmental impact assessment, etc.). Use of toxicity tests in certain government programs is included, with some descriptions of the programs (Section 5). There is introductory coverage of a few more general topics with potential for toxicity testing, such as bioaccumulation, biochemical/physiological changes ("biomarkers"), immune responses and tumour induction, mesocosm studies, or monitoring techniques. Some techniques are still in developmental stages (Section 6).

There are many applications of toxicity tests, although regulatory uses are emphasized here. Toxicity testing can assess effects of toxicants already present in aquatic or terrestrial ecosystems. Testing is essential for evaluation of new commercial chemicals and to develop guidelines and objectives for environmental protection. In concert with chemical analysis, toxicity testing identifies hazardous substances in complex mixtures such as wastewaters, contaminated soil, sediments, or landfill leachates. Toxicity testing can enhance the effectiveness of chemical analyses, reducing costs (Sergy, 1987; Wren *et al.*, 1993). Broad issues are discussed first, beginning with the role of toxicity testing (Sections 1.2 and 1.3), and fundamental concepts and terminology (Section 2). Quality assurance and comparisons of different tests are covered in Section 3. Section 4 discusses data interpretation generally, then proceeds to specific tests in water, sediments, and soil. Section 5 details the roles of toxicity testing in the following Environment Canada programs.

1. Effluent monitoring programs
2. Environmental Effects Monitoring Program
3. Ocean disposal, sediment dredging and monitoring
4. Environmental impact assessments
5. National Contaminated Sites Remediation Program
6. Development of environmental quality guidelines
7. New Substance Notification Regulations
8. Environmental assessment of priority substances
9. Environmental Choice Program

Sources of information. Each document that describes a test method published by Environment Canada contains background information on the test, the species used, and references to supporting details. New or improved methods are continually being developed, however, in environmental toxicology as in all technical and scientific fields. Early developments are often seen in scientific journals such as those listed in the reference section herein. Investigators would be well advised to follow one or more of such journals.

Major sources of useful guidance are becoming available all the time in books, major reports, and compilations on compact disk. Investigators could learn of these by following book review sections or by scanning the lists of environmental offerings from publishers. Some existing books and major reviews are mentioned in appropriate sections of this document. A few others of relevance are given in the following list. General aspects of environmental toxicology are covered in books by Landis and Yu (1998), and Hoffman *et al.* (1995), and more specifically, techniques in aquatic toxicology are described by Ostrander (1997). Compilations of knowledge about toxic substances are often oriented to mammalian effects, but could provide useful background in the library of a laboratory, either in book form (Wexler, 1998) or as a compact disk (ATSDR, 1998). A handbook from Canadian editors provides separate coverage of effects of toxicants on aquatic species (Ramamoorthy and Baddaloo, 1995) and avian and mammalian ones (Ramamoorthy, 1996). OECD (1995a) provides a guidance document for assessing aquatic toxicity, which has broad coverage and many useful formulae, although some sections suffer from brevity, and from uncritical acceptance of certain unique practices developed elsewhere.

1.2 Toxicity Tests Within a Triad of Techniques

Toxicity testing cannot substitute for chemical measurements, or for surveys of communities of organisms. On the contrary, the strengths of toxicity testing are best realized in conjunction with chemical and biological field measurements. These three approaches form a natural triad in which each component enhances the power of the others (Sergy, 1987). Jurisdictions that have embraced toxicity testing to regulate discharges have also emphasized the importance of combining toxicity tests with more conventional chemistry (Blaise *et al.*, 1985; Hunt *et al.*, 1992; Klemm *et al.*, 1990; OECD, 1987). Accordingly, some commentary on the complementary use of (a) toxicity testing, (b) field biomonitoring, and (c) conventional chemical measurements in discharges and receiving environment is provided.

The three techniques in the triad play different roles (see following text). Munkittrick and McCarty (1995) outline the rationale for adopting particular tool(s) including toxicity tests, the quandaries that are inherent in the various approaches, and (unfortunately) the disarray that has sometimes been evident in ecotoxicological work.

Toxicity tests can be regulatory when carried out on a sample of effluent, sediment, or soil, if a limit of toxicity is specified. They can also be used in a predictive manner, by testing the sample and then calculating the degree of toxicity expected in the receiving environment. For effluents, instream toxicity tests are sometimes used (e.g., caged fish in the plume), if it is desired to demonstrate that all modifying environmental factors have been included in the test. Usually it is more efficient to test the much more concentrated effluent or other source of pollution, and use measured dilution factors to calculate the expected toxicity in the receiving environment.

Chemical tests can be used in several ways. First they can be regulatory if a limit is specified for a certain substance in the effluent. Second they can be predictive by calculating from the discharge to concentrations expected in the environment. Third they can be used for after-the-fact monitoring in the environment. The second and third purposes of chemistry are best achieved by comparing with established quality guidelines for water, sediment, or soil.

Biological surveys in the receiving water, or in the substrate of the receiving environment, have obvious importance. Such surveys are *the definitive check* on whether an ecosystem has been affected. That use is a "follow-up" audit, but biotic surveys can never serve in a predictive mode to prevent damage, as can toxicity tests.

Chemistry as a regulatory tool continues to be used in a traditional way to assess the quality of industrial and municipal wastes. Limits of concentration can be imposed for a substance in the waste, or limits on *loading* can be applied (concentration in the waste multiplied by volume of discharge). The Canadian government used both types of limits in recent regulations for pulp and paper mills. For example there were specified maximum concentrations for dioxins, and maximum loadings (kg/d for each tonne of pulp produced) for suspended solids (EC, 1992h; DFO, 1992). Such limits have customarily been tied to "best-treatment" performance of waste treatment plants, variously called "best available treatment" (BAT), "best available treatment economically achievable" (BATEA), etc. Such limits represent good practice in waste treatment engineering, and require the same quality of effluent for a given type of industry, regardless of location or the amount of dilution available in the receiving environment. [1]

Environmental quality guidelines represent another traditional and continuing approach to pollution evaluation and control, quite different from the "best-treatment" tactic. A *guideline* defines the maximum environmental concentration of an individual substance that is judged to be without significant adverse effect. Methods of developing guidelines for water, sediment, and soil are discussed in Section 5.6. The guidelines can be used to erect quality *standards* (i.e., legal limits) or *objectives* (semi-legal goals) that can be monitored in the environment. Monitoring is retrospective, but guidelines can also be used in a predictive, preventive, or regulatory way, by calculating discharge limits that would satisfy objectives after mixing in the receiving environment. Water quality guidelines received wide, almost exclusive use in the USA in the 1970s, and quality objectives have been used by provinces in Canada. Ontario, for example, has had published guidelines and objectives for a quarter-century (OWRC, 1970), with continuing updates. Environment Canada has not made extensive use of quality objectives for water, although a comprehensive and internationally used set of Canadian guidelines exists for water (CCREM, 1987). Quality objectives for soils and sediments are frequently used for contaminated sites (e.g., CCME, 1996a).

The chemical-specific approach has some major advantages. Chemical measurements can be quick, and some are cheap, allowing effective control or monitoring efforts. Once the harmful substances are identified, efforts can be focused to reduce or eliminate their discharge.

Some limitations of the chemical approach follow.

- Periodic sampling might miss peak discharges such as spills, which might be rare but of primary importance to the receiving ecosystem.
- Failure to carry out tests for the most important substances, if it is not known what they are.
- Difficulty in predicting the threat of a complex effluent, based on individual chemical components (Sergy, 1987; Hunt *et al.*, 1992).
- Difficulty in measuring some organic compounds; the proportion readily identified in surface waters represents <10% of the total organic carbon (Galassi *et al.*, 1992).
- The total amount of a substance might be measured, rather than the bioavailable portion of concern.
- A measured concentration might be of uncertain significance because of an inadequate background of toxicological data.
- Cost might be high for measuring numerous substances in a complex effluent, or for some individual substances (dioxins). It becomes impractical to test for everything,

---

[1] A strong environmental case can be made for limiting the *absolute amounts discharged (loading)* for dangerous, persistent, bioaccumulative toxicants such as PCBs or dioxins. They accumulate in food chains, travel long distances (e.g., to arctic ecosystems), and might continue to cause problems for decades or centuries. Therefore control should be imposed on the amounts generated and discharged. The dilution immediately available upon discharge is of little relevance, nor is the actual concentration in the waste discharge an important factor, but rather the weight/amount of toxicant discharged. For more ordinary or "natural" characteristics of a waste such as temperature or oxygen demand, the available dilution could play a part in the control process. Those characteristics carry no connotations of being bioaccumulative or persistent toxicants, and would be stabilized by natural processes in the environment.

and indeed tests might not be available for some substances.

Biological quality objectives could be developed and used in similar fashion to chemical objectives. Indeed they have been used effectively for a decade in USEPA regulations (USEPA, 1984a) and continue to be enhanced (USEPA, 1991a). An obvious objective would be absence of sublethal effects in the receiving water, best ascertained by specified sublethal tests on the effluent, then calculating to the degree of toxicity in the receiving water. The United Kingdom National Rivers Authority proposes to use biotic indices not only for monitoring river quality, but also as quality guidelines and objectives (Section 5.5.4).

1.3 Single-species Tests and their Validity

The toxicity of contaminants can be measured by a number of methods involving one, several, or many species in the laboratory or in the field. The standard test procedures published by Environment Canada are for individual species from freshwater, marine, and terrestrial habitats. They include both lethal and sublethal assessments. The single-species tests have practical advantages of being standardized and generally straightforward, with reasonable economy, precision, and opportunity for replication.

Single-species tests that include ecologically meaningful endpoints such as survival, growth or reproduction, can be used to estimate potential toxic effects in the field. Research comparing the effects predicted from laboratory tests, with actual in-stream effects of complex effluents, has generally found reasonable agreement or somewhat less effect in the field than might be expected (see Section 4.1.1 or review by Sprague, 1997).

A number of more sophisticated within-organism or multi-species techniques might seem more meaningful and useful for assessing pollutant discharges. These are discussed in Section 6 which shows that some are useful in special roles, but most entail difficulties of procedure or interpretation as general-purpose tests. As an example, one such approach is discussed in the following paragraph.

Measuring changes in ecosystem functions sounds like a broad and meaningful approach for assessing pollutants. Effective environmental protection must encompass not only individual species, but entire ecosystems and their processes such as primary production, decomposition, and nutrient cycling. Therefore, assays such as primary production or soil respiration might seem superior to extrapolations based on single-species tests. However, while numerous process-oriented assays are available from the research community, many are relatively undeveloped or lack standardization. Furthermore, where comparisons have been made, species-level toxicity has invariably been a more sensitive measurement of stress than ecosystem functions such as respiration or production, at least in aquatic systems (Schindler, 1987). Because of functional redundancy and resilience within the biological community, ecosystems evidently respond to stress by maintaining essential functions at the expense of sensitive species [2]. A community that begins to show a decline in functions has already suffered major damage to its species composition, and exhausted its capacity to deflect stress through species replacement or other mechanisms. Sensitive species or populations would thus show toxic effects at lower concentrations than would ecosystem functions. Accordingly, single-species testing allows early remedial or preventive action.

1.4 Benefits and Limitations of Toxicological Approaches

Benefit no. 1 in Table 1 far outweighs all others because biological tests are the *only* way to measure toxicity; chemical measurements can predict but can never measure toxicity. Conversely, *limitation no. 1*, that the biological test cannot identify the substance causing an effect, is the strongest disadvantage. These two sides of the coin are why it might be important to use both biology and chemistry as complementary approaches. Used together in a toxicity identification process, they can sometimes identify toxic components, if that is a desired objective.[3] *Benefit no. 2* (Table 1) is a special case of no. 1, because biological tests, properly done, *always* measure toxicity faithfully, even for "mystery pollutants" (but see discussion of accuracy and precision in Section 2.6).

---

[2] For example, there might be several predators with a similar range of food species, and functions like decomposition or primary production would reflect the contributions of many species. Loss of a few species in any category, because of toxicity or other environmental conditions, would likely be compensated by increased numbers or activity of remaining species. The degree of predation or overall rate of a function might show little change (Cairns, 1983). Primary and secondary productivity are surprisingly stable under all but the most extreme conditions of environmental deterioration (Reice and Wohlenberg, 1993). Using ecosystem functions as toxicological endpoints remains an area of active research; Cairns *et al.* (1992) and Reice and Wohlenberg (1993) review the topic for aquatic systems.

Benefit no. 3 is the strong attraction of obtaining an early sign of potential effects on biota, so that preventative measures are possible. With such a prediction, a field survey could be carried out to assess actual effects on populations, or a chemical investigation could attempt to discover the cause of toxicity. This benefit is restricted somewhat by *limitation no. 2*, a possible slowness in obtaining an answer, e.g., four days in a test with fish[4]. Chemical tests can measure some toxic chemicals or conditions quickly, frequently, and cheaply, or even monitor continuously and automatically (e.g., low oxygen, pH, certain toxic molecules).

Benefit no. 4 is that toxicity tests give an integrated answer to the total effect of diverse toxicants in a given complex effluent. There are chemical methods for predicting toxic effects of a mixture, but they remain predictions, and might not forecast the actual toxicity in the receiving environment (reviewed in Broderius, 1991; Christensen and Chen, 1991; Sprague, 1992).

Benefit no. 5 is that results expressed in terms of harm to living organisms are often more credible or comprehensible than chemical data, to non-specialists (Sergy, 1987). *Limitation no. 7* is the counterbalance that a "toxic" finding could be misleading if used in simplistic ways, e.g., failing to allow for dilution or mitigating factors.

Benefit no. 6 indicates that single-species testing can be considerably less expensive than comprehensive chemical analysis. However, some toxicity studies can be expensive. Those investigating populations, communities, or ecosystems, for example, could require many different single-species tests, multi-species microcosms or mesocosms. Clearly, these tests should be used only when necessary for the problem under consideration (Sergy, 1987). The purpose of a test and the kind of information expected, therefore, must be thoroughly understood before tests are selected.

Limitation no. 3 is true, that only a few species are used in the laboratory. It would be prohibitively expensive and logistically overwhelming to attempt to test even a small proportion of the species in a given natural community. Standard or surrogate species are deliberately selected for sensitivity, among other features; the implicit assumption is that they represent sensitive species found in nature. That assumption finds general support in field validation of laboratory tests (Section 4.1), so limitation no. 3 is apparently minor, as long as diverse types of organisms are tested (e.g., a fish, an invertebrate, an alga, or a bacterium).

Limitations no. 4 and 5 are also true, that laboratory tests do not normally include the natural environmental stressors, nor attempt to mimic the natural variability, of the environment. The tests are standardized for good reasons. It is intended that organisms should be tested under optimal conditions, without confounding effects of other stressors such as low dissolved oxygen or temperature extremes, which might contribute to a false or exaggerated indication of toxicity. Tests attempt to follow principles of good experimental design by controlling all variables except the one of concern, the toxicant. Ancillary variables (e.g., temperature, type of water) are constant and favourable, and the toxicant is varied by a series of fixed concentrations. If desired, a series of fixed ancillary conditions could be added (e.g., different fixed values of pH or temperature). The fixed levels allow effective methods of analyzing results and definitive interpretations, which can then be used to predict to the real world with its fluctuation in concentrations and its spectrum of modifying environmental conditions. As an associated item, an

---

[3] Sometimes an investigator might not wish to identify the toxic substances; it might be enough to know if there were toxicity, in order to take appropriate action. Some complex industrial effluents might have thousands of potential toxicants, and it could be nearly impossible to identify them all, let alone quantify individual toxic contributions (e.g., effluents from pulp mills).

[4] Limitation no. 2 is a qualified one. Most new toxicity tests are fast and some yield results in as little as half an hour (e.g., modern tests with microorganisms, protozoans, echinoids, and rotifers).

Table 1 **Benefits and Limitations of Toxicity Testing**

Benefits

1. Provides a direct estimate of lethal or sublethal effects. Chronic tests approximate long-term effects of toxic materials.

2. Measures toxicity when the toxic agent is not known, not identified chemically, or lacks toxicological data.

3. Can provide an early warning or prediction of potential environmental damage.

4. Accounts for effects of toxicant mixtures. A chemically complex effluent can be assessed as a single generic pollutant.

5. Allows results to be more immediately understandable, credible, and convincing to industry or the public.

6. Is usually less costly than chemical analyses. Toxicity tests become more cost-effective than chemistry with increasing complexity of the matrix of pollutants.

Limitations

1. The substance causing toxicity is not identified (although chemical/biological Toxicity Identification procedures can assist in this).

2. Traditionally, has been relatively slow compared to simple chemical analyses. However, newer toxicity tests are quite rapid.

3. Only a few species are tested, out of many present in ecosystems.

4. Test organisms are not exposed to other stressors (e.g., extremes in temperature or water current, predators, disease vectors) which would be present in the wild. Conversely, test organisms do not benefit from natural sorption, detoxifying, or other mitigating mechanisms which might be present in nature. Nor do organisms have the opportunity to avoid the toxicant.

5. Variability of natural environment is not represented.

6. Normally, does not directly measure bioaccumulation or retention of chemicals. However, there are methods combining toxicity and bioaccumulation with molluscs and other organisms.

7. Simplistic use of results might mislead public or regulators.

---

array of tests on different species greatly increases the information and its usefulness for predictions.

Limitation no. 6 indicates that toxicity tests normally measure only immediate bioavailability of the toxicant(s). They cannot measure the degree of bioaccumulation or retention of a chemical; for that, separate methods must be used (Klemm *et al.*, 1990). (In developing quality guidelines for water, soil, or sediment, or in ecological risk assessments, the separate data for bioaccumulation are normally integrated into the process.) Similarly, toxicity tests on an effluent before discharge cannot account for chemical or physical changes after it enters the environment, with a possible increase or decrease in toxicity (Sergy, 1987). Such issues can be resolved, at least partly and *a posteriori*, by directly testing the toxicity of affected receiving waters, soils, or sediments.

Precision of toxicity tests is not listed as a topic in Table 1 because it is generally similar to precision of chemical tests. For both lethal and sublethal tests, recent comparisons within and among laboratories have found coefficients of variation that are commonly in the 30–50% range. That variation is appreciable, but is comparable with variation of chemical measurements, especially for organic chemicals and for inorganics near the detection limit (see Section 2.6).

1.5 Support for Single-species Tests

Canada was an early leader in the regulatory use of toxicity tests. Some provinces, notably Alberta and Ontario, started using toxicity testing for regulatory purposes in the 1950s. A quarter of a century ago, Environment Canada promulgated regulations and guidelines requiring that effluents from new pulp mills be non-lethal to rainbow trout (EC, 1971; 1972)[5].

Six years later, T.E. Howard of B.C. Research stated that "no other country demands that industry satisfies a biological criterion of acceptable effluent quality", and that "the existence of such standards has required newer (pulp and paper) mills to install secondary treatment facilities" (Howard, 1978). Environment Canada concluded that a series of single-species toxicity tests, used judiciously, can improve environmental protection (Sergy, 1987).

Over the years, government, industries, and international agencies have increasingly adopted single-species toxicity tests to predict potential toxic effects in the wild, and endorsement has been particularly striking in the last decade. Ontario incorporated toxicity testing into a major program for regulating discharges the Municipal and Industrial Strategy for Abatement (MISA) (see Bonsor *et al.*, 1988). Single-species testing (lethal and sublethal) became part of the control strategy of the USEPA in the mid-1980s, and now receives heavy use in granting discharge permits (Wall and Hanmer, 1987; USEPA, 1991a; 1994a). Regulatory toxicity testing with single species is used throughout western Europe including the U.K. (Hunt *et al.*, 1992).

The American Society for Testing and Materials (ASTM) has developed many standard methods for ecotoxicology. A method for fish and other aquatic animals, for example, was available in 1980 and is still current (ASTM, 1997b). As well, the International Organization for Standardization (ISO) has been producing standard methods since the early 1980s. Another active agency is the Organization for Economic Co-operation and Development (OECD), of which Canada is a member. The OECD concluded that toxicity testing is a scientifically sound tool that should be used in member countries, in conjunction with chemical and ecological measurements, to identify, quantify, and control the discharge of toxic pollutants (OECD, 1987). Working groups have published many standard biological methods (e.g., OECD, 1992j). The OECD has been endeavouring to develop standardized approaches and sets of tests for testing industrial chemicals. The initial attempts of a working group to provide synoptic comments and tabular evaluations for a wide variety of tests are embodied in a large report (OECD, 1998b). The complexity of this stage of development means that the report offers little specific guidance to Canadian investigators at this time. The European Economic Community (EEC) is also beginning to promulgate ecotoxicological tests (EEC, 1992a, b). While these examples are drawn from effluent testing, the usefulness of combined chemical and toxicological testing applies equally well to soils and freshwater or marine sediments (Wren *et al.*, 1993; ASTM, 1997a).

The government of Canada has proposed a policy "to virtually eliminate from the environment substances that are predominantly anthropogenic .... persistent, bioaccumulative and toxic" (EC, 1994a). [6] A substance is to be considered toxic if it might enter the environment in a quantity or under conditions that "have or may have an immediate or long-term harmful effect on the environment" (*ibid.*). Biological tests are obviously required for the fulfilment of that policy. Toxicity testing is already required under regulations of the Fisheries Act, the Canadian Environmental Protection Act (CEPA), the Transportation of Dangerous Goods Act, the Pest Control Products Act and other acts of the Canadian federal government.

---

[5] The regulations and guidelines outlined a simple test that had many good features for monitoring. The definitive test for legal purposes required ≥80% survival of rainbow trout in 65% effluent. Continuous flow of test solution was recommended, with replicates. If fewer than 80% survived, the effluent was considered "toxic" and discharge was an offence under Section 33 of the Canada Fisheries Act. An alternative monitoring test could be static, although daily change of test water was recommended.

---

[6] Clearly and appropriately, the policy is aimed at those man-made organic compounds which have caused such concern, for example PCBs, persistent pesticides, and toxic dioxins.

Basic Concepts of Environmental Toxicology

This section reviews some essential terms and concepts of toxicity testing, and can be used in conjunction with the glossary. Particular reference is given to the standard test methods published by Environment Canada, each of which provides some relevant background rationale. More extensive information on ecotoxicology can be found in Rand and Petrocelli (1985) and Rand (1995); the earlier reference being possibly more useful for basic information.

2.1 Terminology

Key Guidance

- *Toxicity is the inherent capacity of a material to cause adverse effects on living organisms. Toxicity can be lethal or sublethal (e.g., an effect on growth); either of those categories might be acute (rapid) or chronic (long duration).*

- *All tests are characterized by an organism, a medium, and an endpoint. Each has a number of options that should be given careful consideration in design. A test might be multi- or single-concentration, static or renewed, have one or two kinds of controls, and nominal or measured concentrations.*

The test methods published by Environment Canada define toxicity as *the inherent potential or capacity of a substance to cause adverse effects on living organisms*. Section 11 of the Canadian Environmental Protection Act (CEPA, 1988) defines a toxic material as... "A substance is toxic if it is entering or may enter the environment in a quantity or concentration or under conditions (a) having or that may have an immediate or long-term harmful effect on the environment; (b) constituting or that may constitute a danger to the environment on which human life depends; or (c) constituting or that may constitute a danger to human life or health." (EC and HWC, 1993).

Adverse (toxic) effects include death of the organism (*lethal* toxicity) or some *sublethal* effect such as impairment of growth. Each toxicity test is designed to measure one or more of these adverse effects, termed *endpoints* of the test. *Acute* toxicity occurs rapidly, relative to the lifespan of the organism, during or after a short exposure of minutes, hours, or days. *Chronic* toxicity becomes evident after a longer period of exposure, usually 10% or more of the organism's life span. For example, the most commonly used acute exposure for fish and larger aquatic invertebrates is four days; if such an organism had a normal life-span of two years, a chronic exposure would be in the order of two months or longer[7].

Death is the most commonly measured acute effect. Chronic lethality is sometimes measured, but most chronic tests measure more subtle effects such as reproductive impairment. Many sublethal effects such as inhibition of growth, or metabolic changes, could be measured in either a chronic exposure or an acute one such as 7-d exposure of fish. Acute sublethal effects such as attraction or avoidance responses or rapid biochemical alterations signifying stress responses, can be measured during or after brief exposures of minutes, hours, or a few days, at concentrations that would not cause death in conventional lethal tests. The most complete toxicity test is the chronic exposure of a *life-cycle test*, in which organisms are exposed to the

---

[7] Acute and chroinc are relative terms. Because different species have life spans of widely varying duration, an acute exposure for one organism might be a chronic exposure for another. For example, an acute exposure of four days for a vascular plant would not only be chronic for a bacterium or an alga, but it might span several of their generations. One should be aware when reading reports and publications in aquatic toxicology, that some authors apparently consider "acute" and "lethal" to have the same meaning, and others mistakenly write "chronic" when they mean "sublethal". This confusion appears to have arisen because many familiar acute toxicity tests use lethality as the endpoint, and because chronic tests almost always measure sublethal toxicity. Unfortunately, such errors are widespread, even in publications of the USEPA (e.g., Klemm *et al.*, 1990), and contribute uncertainty to publications in environmental toxicology. Care should be taken to use the terms correctly. The terms "sub-acute" and "sub-chronic" are also seen in the literature, but they remain without precise meaning and are probably best avoided.

test material for a full generation, from, for example, adult through gamete formation, fertilization, development of offspring and adult again. In such a test, observations can be made on a spectrum of sublethal effects such as reproductive success, growth, tissue abnormalities, and physiological variables. For reasons of practicality and brevity, most sublethal tests concentrate on a relatively short exposure using specific sensitive segments of the life cycle, especially the developmental life stages (e.g., the 7-d, 30-d, or 70-d test options with early life stages of rainbow trout (EC, 1998).

Toxicity tests are sometimes referred to as biological *assays* or *bioassays*, because they use living organisms to measure the potency or strength of materials. However, while all toxicity tests are bioassays, some bioassays are not toxicity tests (Chapman, 1989a). The more explicit term *toxicity test* is standard terminology in all Environment Canada test methods. Herein, "toxicity test" refers to single-species tests using laboratory procedures, unless stated otherwise.

2.2 Common Features and Design of Tests

All environmental toxicity tests have three characteristics: the organism, the medium, and the endpoint. Most major groups of organisms are now represented by standard test methods; this includes bacteria, protozoa, nematodes, algae, vascular plants, insects and other arthropods, molluscs, annelids, amphibians, and fish (see Section 2.3). The medium can be fresh or saline water, aqueous extract from soil or sludge, aquatic sediment, or terrestrial soil (Section 2.4). The endpoint is assessed by comparing effects caused by the test material to effects found (or not found) among *control* organisms subjected to identical conditions except for absence of the test material. Although toxic effects are adverse by definition, low concentrations of a material can occasionally stimulate growth or some other biological function, indicating that the material is biologically active although the conventional toxic effects are not observed.

The medium for diluting the test material is usually specified in methods documents. Normally, the medium is the laboratory water, clean soil, or sediment that is used for maintaining laboratory cultures, or an artificially prepared medium known to be suitable for the organism. The diluent medium (water, sediment, or soil) could substantially modify the outcome if toxicity were modified by such variables as pH, water hardness, adsorption onto particles, changed bioavailability, or if the particle size of the sediment affected the organisms (see Section 4.15). Therefore, the diluent must be chosen judiciously, with a full appreciation of the kind of information the test is expected to provide. Tests intended to predict effects in a particular location might require a diluent drawn from that environment, such as water or sediments taken upstream of a wastewater discharge or an unaffected reference site [8].

Endpoints differ between two major categories of tests. In the commonly used *multi-concentration test*, a defined number of individuals of the selected species is exposed in each of several test vessels. The vessels contain a series of concentrations of the test material, normally in a geometric dilution sequence (i.e., a logarithmic series) [9]. The exposure uses specified constant conditions and a fixed amount of time. The water, soil, or sediment used in the controls and for dilution, must be from a source known to support the test organism without detrimental effect. Results can be used to make a quantitative estimate of the strength or degree of toxicity of the test material, i.e., estimates define a *concentration* such as the LC50, EC50, or IC25.

---

[8] Similarly, the control might be water flowing into part of an industrial process, and that same water would be used for making dilutions for the test solutions. In unusual circumstances, there could conceivably be a series of concentrations of both the upstream "control" water and the downstream water, diluted by clean laboratory water; the severity of effect could be compared.

[9] In a logarithmic series of concentrations, each level is greater than the preceding one by a constant proportion. That constant increase across the range of tested concentrations can be appreciated at a common-sense level, as being desirable for a balanced test. On the other hand, the unsuitability of a fixed arithmetic increase can be seen in some simple examples. One could possibly run a successful test with the arithmetic series 1, 2, 3, 4, and 5 mg/L. Nobody, however, would be likely to set up test concentrations of 1001, 1002, 1003, 1004, and 1005 mg/L, because the increase in concentration is such a feeble change. There is a question of *proportion*, and to give the high-concentration test a series parallel to the low-concentration test would mean 1000, 2000, 3000, etc. The same logic applies to maintaining a proportional change within a single, narrower series. The change from 1 to 2 mg/L is a doubling, or 100% increase, while the change from 4 to 5 mg/L is only a 25% increase in stimulus. Accordingly, tests are run with a geometric/logarithmic series such as 1, 2, 4, 8, and 16 mg/L, each concentration presenting its test organisms with an exposure twice as great as the previous lower one.

In the alternative *single-concentration test*, the effect observed in the single concentration is compared with performance in the controls. The endpoint is simply the presence or absence of toxicity, and there is no quantitative estimate of the strength of toxicity. Such tests are usually designed carefully to be simple, economical, and clear-cut regulatory tests, and are sometimes called pass/fail tests. They often use full-strength samples of wastewater or contaminated soil or sediment.

Controls (sometimes called *negative controls*) are a key element for determining endpoints in all kinds of toxicity tests, since they provide the baseline of biological functioning in the absence of the test material. Controls also serve as general monitors of quality in the experiment. For example, a sudden temperature change, handling stress, or unnoticed disease might cause mortality, even in the absence of a toxicant. Similarly, soil that is too acid or too wet to support earthworms will cause mortality or growth inhibition, even without adding a test material. Standard toxicity tests all include criteria for the maximum acceptable level of biological effect in negative controls, if a test is to be considered valid (see Section 3). In some sediment tests, it is common to run double controls, one being uncontaminated samples from the area where test sediments were collected (the *reference sediment*) and the other being sediment from the collection site for organisms or else standard laboratory sediment.

Tests with *reference toxicants*, sometimes called *positive controls*, represent another way of detecting changes in performance of organisms and evaluating precision of tests. A reference toxicant is a pure chemical, commonly a metal (e.g., zinc) or a simple organic compound (e.g., phenol), having a well-documented degree of toxicity to the test species. Multi-concentration tests with the reference toxicant are used to detect effects outside the normal range. Unusual effects would be taken as an indication of changes in the organisms, possibly a poor state of health, or changes in the medium that modified toxicity of the reference chemical. Both positive and negative controls are needed to assess precision of tests, and requirements in that regard are given in all Environment Canada test methods as well as in specific guidance manuals (EC, 1990d; 1995a).

Toxicity tests can be categorized according to the type of renewal of test material. In *static* tests, organisms are exposed to one or more concentrations of the test samples in isolated vessels (e.g., beakers, buckets, aquaria) and the test material is not renewed for the duration of the test. *Flow-through* tests can be used for fluids; the solution is renewed more or less continuously in each vessel for the duration of the test. *Static-renewal* tests represent a compromise, with the test material partially or wholly renewed in a batch method, usually daily. That would involve replacement of the solution in an aqueous test, overlying water in a test with sediment, or soil mixture in terrestrial tests. The test methods published by Environment Canada give the requirements for validity of any toxicity test, in whichever mode or modes is/are specified for the test [10].

The concentrations of toxicant in test vessels can be stated as the "added" or *nominal* (from dilution calculations), or the *measured* concentrations obtained by direct chemical measurements on aliquots representing the contents of test vessels. This is another way of categorizing toxicity tests into two classes, and it should be clear which class applies to any given test [11].

---

[10] Each mode of testing has specific advantages and detriments, and each mode can be well- or poorly designed. Static tests are more likely to suffer a decline in toxicant through physical or biological decay, loss to the atmosphere, assimilation by the test organisms, and/or adsorption onto the walls of the container. Sometimes when testing large organisms it is difficult to have the required amount of test material (and large enough tanks), but compromising the amounts might render the tests meaningless. Static or static-renewal tests are the method of choice for testing aquatic sediments, and can provide meaningful and convenient monitoring of waste discharges, for which they are often used. Flow-through tests are particularly useful for research-level investigations of pure chemicals or volatile substances; they cannot be used for soils or sediments, nor for some fluids which settle or clog the apparatus. The volume of test solution and amount of inflow must be carefully calculated, or else the flow-through test might become essentially static. As a rule of thumb, the inflowing volume must equal three times the standing volume to achieve 95% molecular replacement in a given time period. In static-renewal tests, the renewal can be difficult to carry out in some instances without disturbing the test organisms unduly and possibly affecting the results.

2.3 Organisms Used in Tests

Key Guidance

- *Standard test species are chosen for their adaptability to laboratory conditions, and also as representatives of sensitive species in the environment. More methods are becoming available for smaller organisms which have reasonably short life cycles, allowing tests of reproductive performance within a reasonable time.*
- *Tests with standard species provide a uniform base of information.*
- *Species and test should be chosen to provide toxicological information that is relevant to the ecosystem of concern.*
- *A first priority is to test diverse types of organisms, using a battery of standardized methods and associated organisms (e.g., microbes, plants, terrestrial and/or aquatic invertebrates, and fish. Beyond that, suitable local species might be used if site-specific information were required, but parallel tests with a standard species should also be done.*

The standard species used in routine programs of toxicity testing are limited in number but now represent a broad spectrum of *types* of organisms. One or two sensitive standard species can be used in repetitive monitoring of potential pollution, while a battery of tests with diverse kinds of organisms is best for initial investigation of a new chemical or source of potential pollution. While limitations on species suitable for laboratory tests have to some extent restricted the list of available species, the customary widespread use of a few standard tests is desirable for comparing toxicities of various materials.

2.3.1 Desirable Features of Test Species

Organisms for laboratory tests should possess a number of features. They should be relatively easy to handle and maintain in the laboratory, and should remain healthy. The organisms should be fairly easy to enumerate in tests, exhibiting a clearly defined and measurable endpoint.

For chronic tests, there is a strong trend toward using animals with short, simple life cycles so that reproductive effects can be included in a reasonably short test. The water fleas *Ceriodaphnia dubia* and *Daphnia magna* are widely used for tests, because they mature and reproduce in a few days, and about a week, respectively. It is easy to count their offspring which are produced parthenogenetically. Absence of sexual reproduction in the test removes it as a confounding factor, but could preclude assessment of some reproductive effects such as endocrine disruption.

If the test species or a near relative is common and widely distributed in Canada, test results will probably seem more relevant to most people. From a scientific or technical viewpoint, a species need be neither common nor native (see Section 2.3.2). It should be similar in sensitivity to native species, but broad similarities do exist within groups such as fish, crustaceans, etc., so that substitutions can be appropriate. It is important that each test species should give dependable and sensitive performance in the laboratory, and it is important to include a wide variety of organisms in a test program. Both those factors should be satisfied when choosing the test species. If there is any doubt that the selected species might fail to represent sensitive species found in the local habitat, then local representatives could also be tested.

The three origins for test organisms are culturing in the laboratory, rearing, or collecting in the field. *Cultured* species can be maintained indefinitely in a laboratory through many generations (e.g., fathead minnows), or in commercial cultivation (e.g., lesser duckweed and lettuce). Cultured species can usually be obtained from biological supply houses, university culture collections, or government agencies. They often represent only one or a few clones or strains of the species, and therefore have little genetic variation (Soares *et al.*, 1992). That is both advantageous (less variability in results) and unfortunate (less representative of wild species). *Rearing* an organism implies collecting young from the wild and maintaining them in the laboratory until they are the right size or age, or until healthy offspring are produced for use in a test. Finally, *collecting* wild organisms without necessarily maintaining them in the laboratory for a significant length of time, might be satisfactory for some tests. Environment Canada uses field-collected organisms for its acute toxicity test with sticklebacks and marine or estuarine amphipods, and for a sublethal test of fertilization success in sea urchins or sand dollars. Field-collection might also be used when it is important to test a particular indigenous species. The organisms should be acclimated to the appropriate conditions of temperature, salinity, and lighting before being used in tests.

---

[11] When testing complex mixtures such as effluents or sediments, there is little choice but to use nominal concentrations, although some component of the test material might be chemically measured as a check of concentrations. Nominal concentrations are clearly inferior when testing a chemical, and might give a totally unrealistic depiction of what the organisms were exposed to. If chemical measurements of concentration are made, they should be used in calculating the endpoint (LC50, IC25, etc.), and the nominal or intended values can be set aside. The chemical measurements must be accompanied by an evaluation of the accuracy and precision of the chemical method used and the particular measurements.

2.3.2 Sensitivity of Test Organisms

Sensitivity to a range of toxicants is a key attribute of a test species because one of the main assumptions of toxicity testing is that concentrations shown to be without harmful effect in the laboratory, will also be protective of natural biological communities including sensitive and less sensitive species. Over recent decades, there has been a major process of sifting out sensitive species for use in the laboratory, and appreciable amounts of effort have compared their sensitivities to various classes of toxic substances (LeBlanc, 1984; Schimmel *et al.*, 1989).

While it is important to pick sensitive standard species for the laboratory, differences in sensitivity might not be large among most members of a given group (e.g., fish). The really large differences in sensitivity are found among groups (e.g., differences between fish and crustaceans, and between crustaceans and bacteria). Furthermore, no particular group is most sensitive to all categories of toxicants. There is abundant evidence that there is much greater variation in sensitivity among invertebrates, algae, and other organisms (*as groups* as well as individually), than is found among fish [12]. Therefore, adequate exploratory testing of a chemical's toxicity should include various organisms from different trophic levels and taxonomic groups, rather than different species of one group such as fish. Using such a *battery* of toxicity tests is important for assessing "new" toxicants. It is also especially important for assessing complex mixtures like municipal and industrial effluents, and contaminated sediments or soils, because different types of organisms might be affected by entirely different categories of toxicants in the test substance (van Straalen and Denneman, 1991; USEPA, 1991b).

---

[12] The variation in sensitivity *within* a group of similar organisms is not great, and fish can be taken as an example group that has been well studied. For many toxicants, especially those with generalized modes of action, the variation in tolerance by different species of fish is often within five-fold or at most ten-fold, which might seem large but is actually about the same as is found for within- and between-laboratory differences in test results for a single species of fish (reviewed in Sprague, 1985). As an example of similar sensitivity of various fish, 10 organic chemicals including those which elicited at least four modes of action were tested for lethality to six species of fish by Thurston *et al.* (1985). Nine of the chemicals had low ratios of 1.5 to 4.6 between the LC50s of the most tolerant and most sensitive species. For the 10th chemical (the insecticide Permethrin) goldfish were 77 times more tolerant than the most sensitive species, an appreciable difference but not large compared to differences among types of organisms mentioned below. There was no strong evidence of a "most sensitive" or "most tolerant" fish. Goldfish were most tolerant in three of the 10 comparisons, mosquitofish and bluegill in two cases each, catfish and rainbow trout in one case each, and fathead minnows in none of the comparisons. Similarly, Fogels and Sprague (1977) concluded that the small tropical flagfish and zebrafish were not greatly different from rainbow trout in their lethal tolerance, compared to usual between-laboratory and within-laboratory differences. Irwin (1965) tested 57 species of fish in an oil refinery waste and found only a five-fold range between the most sensitive (gizzard shad) and least sensitive species (guppy). [Unusual major differences in sensitivity among fish are the four-fold to 900-fold ranges for various organophosphate insecticides (Pickering *et al.*, 1962)].

There is abundant evidence that variation in sensitivity *between* groups is much greater, and an example would be comparisons among invertebrates, algae, and other organisms, and the fish previously described. Stephan (1982) compared acute lethality of 21 chemicals for 8–45 diverse organisms, in fresh or salt water or both. Altogether he obtained 38 ratios of highest/lowest LC50s; 13 were >1000, and 15 more were >100. LeBlanc (1984) and Schimmel *et al.* (1989) compared sensitivities of fish, invertebrates, and algae to various metals, organic chemicals, and industrial effluents. Effects generally ranged through two or three orders of magnitude for the same test material. The order of sensitivity among organisms was seldom the same for different materials. Cladocerans tended to be sensitive to metals, but relatively resistant to most pesticides, while fish tended to be more sensitive to organics and ammonia.

The principle of using a battery of toxicity tests is widely recognized in recommendations for deriving water quality guidelines (Stephan *et al.*, 1985; Giesy and Hoke, 1989). Some agencies or individuals have concluded that ecotoxicological assessment requires at least three species from different phyla and trophic levels (USEPA, 1991a; OECD, 1992b; EC and HWC, 1993). A battery consisting of an alga, an invertebrate, and a fish is recommended by the USEPA for testing of industrial and municipal effluents (USEPA, 1991a) and by Environment Canada for Environmental Effects Monitoring in the pulp and paper industry (EC, 1997c). From a massive comparison of data, Sloof *et al.* (1983) concluded that screening of toxicity with a fish, daphnid, and green alga often failed to indicate the full extent of toxicity, and recommended that the diatom *Microcystis aeruginosa* should also be included to increase sensitivity.

Once the toxicity of a material has been established using the battery of tests, any future monitoring programs could revert to testing with a single species. This might be the most sensitive organism from the exploratory work, but it need not be. Any standard, sufficiently sensitive species could be used in monitoring, with regulatory standards set suitably for protecting the most sensitive elements of a community. For example, rainbow trout might be used because it is a familiar species and a wide data base exists from other monitoring programs.

2.3.3 Choice of Species for Local Relevance

In selecting species for a toxicity test, there might appear to be two divergent influences, but they are not entirely in conflict. Testing with a standardized species is important, but the species should be relevant to the ecosystem of interest, to provide meaningful information. For most routine uses of toxicity tests, however, the species can be relevant without necessarily being native to the ecosystem. Species recommended for standard methods have been selected for attributes that include sensitivity, and they are intended to represent species in environments similar to their own (USEPA, 1991a). Although there will be some differences in sensitivity from, say, one fish species to another, the really big differences will be among diverse types of organisms (Section 2.3.2). Thus it can be satisfactory to assess a local situation by tests with a non-native but generally relevant fish, but it is important to do initial tests with diverse organisms.

Obviously irrelevant species should not be used, as in the following list of rather extreme examples.

1. species found in entirely different environments (e.g., using a marine amphipod to test effects of freshwater sediments)

2. species so far removed taxonomically and physiologically that results could not be generalized (e.g., using a plant to assess toxicity to animals)

3. species with atypical metabolic requirements

4. species which would be exposed to the toxicant(s) through different routes or to a different degree (e.g., using a pelagic fish to assess sediment toxicity).

The relevancy argument has been levelled against the luminescence assay using a marine bacterium, for assessing toxicity in fresh waters. That is valid, but the positive issue would be including a bacterium in the battery of tests, rather than leaving out micro-organisms for lack of a convenient method.

Local species have been substituted for standard species when site-specific information was required, such as in assessing contaminated sites or setting local water quality objectives. Tests with a local species can also be warranted if it is ecologically or economically important, particularly at risk because of sensitivity to the toxicant, or if its life-style increases its exposure. Some potential difficulties must be considered. The local species should remain healthy in a laboratory setting, and within-laboratory variability should be examined by using reference toxicants. Parallel tests with standard species should be run. Fortunately, the question of relevant local species is becoming less important as more standard tests are developed with new species; such development is an active field.

2.4 Materials Tested

Testing of *environmental samples* such as wastewater, contaminated soil, or sediments, can be distinguished from testing *individual chemicals*. Toxicity of environmental samples is often assessed by the effect of full-strength material, as an alternative to estimating the percent of full strength material that is needed to produce a defined effect such as 50% mortality. The concentration or even the identity of the toxicant(s) producing the effect might remain unknown. By contrast, testing of individual chemicals involves pre-measurement to produce a series of test concentrations, with a toxicity estimate in terms of concentration (e.g., mg/L or mg/kg). Tests on individual chemicals can use any concentration necessary, theoretically, up to the capacity of the test medium to carry the chemical, whether dissolved, adsorbed onto particles or as micelle fractions. For environmental samples, however, the upper limit for concentration is full strength. If that has no effect, the sample can only be characterized as nontoxic.

Testing of soils and sediments is less developed than aquatic testing but is advancing rapidly (Wren *et al.*, 1993). Testing a solid medium such as contaminated soil or sediment is often done by a single-concentration test using 100% sample. For some purposes, clean soil or sediment might be spiked with a range of concentrations of a chemical or mixtures of chemicals (EC, 1996a; USEPA, 1994d). Aqueous samples derived from the solid media might also be tested. *Leachate* is a liquid that has percolated through a column of soil or solid waste, while that solid remains in place in the environment. An *elutriate* is an aqueous solution prepared in the laboratory by adding water to a solid or semi-solid material, shaking, then obtaining the liquid by centrifuging, filtering, or decanting. *Porewater* is the water naturally found in interstices between particles in sediment, and it can be separated out by centrifuging, squeezing, or filtering (EC, 1994c).

For soil and sediment samples, concentrations of potential toxicants can be higher in small-volume elutriates than in leachates because of greater interaction during extraction. Differences related to technique are not trivial, and methods continue to develop. Recent techniques approved by Environment Canada and others should be followed. Aqueous extracts, once obtained, can be tested using variants of standard aquatic methods.

2.5 Estimation of Endpoints from Concentration and Effect

Key Guidance

- *In a **quantal** test, each organism either shows the effect or does not. The endpoint is the concentration causing an effect in the median organism, the LC50, or the EC50 (see glossary).*
- *A quantal endpoint is estimated from a line fitted over the range from low to high effect, usually the probit of effect plotted on the logarithm of concentration. Greatest precision is usually obtained with the median response.*
- *In a **quantitative** test, each organism can show a graded effect, such as size attained or number of eggs produced.*
- *The preferred quantitative endpoint is an estimate of a point on the concentration-response distribution, the ICp, usually the IC25. Improved methods are needed for estimating the ICp and its confidence limits. Another category of endpoints comprises the NOEC, LOEC, and TOEC. These are less desirable because their values are influenced by the design and quality of the test.*

In ecotoxicology, there is usually particular interest in the concentration of toxicant in the water, soil, or other medium that surrounds the organism. The *dose* of toxicant within the organism is often unknown, unlike classical mammalian toxicology in which the dose is all-important, and is obtained from the diet, injection, inhalation, or by other means.

Guidance on statistical methods for determining endpoints of toxicity tests is provided in each test method published by Environment Canada's and in the guidance manual on statistics (EC, in prep., a).

2.5.1 Quantal Tests

The example of a *concentration-effect curve* (or concentration-response curve) in Figure 1 represents results from an aquatic toxicity test. Figure 1 relates the intensity of the biological effect to the concentration in the test medium, and represents a lethality test; this is a

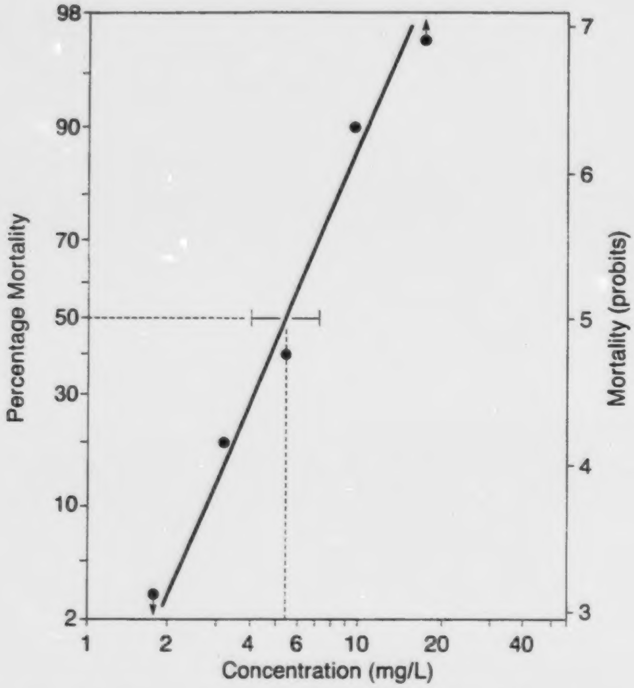

Figure 1 **Results of a Lethal Test Plotted on Logarithmic-probability Paper**. A straight line has been fitted to the data. The LC50 has been estimated graphically as about 5.5 mg/L by the line drawn across from 50% effect, then vertically down from the intersection with the fitted line. Probit analysis on computer confirmed the LC50 as 5.6% and estimated the 95% confidence limits shown by the horizontal bar at 50% effect.

quantal test (all-or-none) in which each organism either shows the effect, or does not. The tested concentrations should produce several partial effects near the middle of the percent scale, since such values have the least variability and the most value or "weight" in estimating the nearby endpoint for 50% effect. There should also be a pronounced gradient of effect, ranging from little or no effect at a low concentration to complete or nearly complete effect at a high concentration. Those very low and very high percent effects have reduced weight in fitting the line, but they help to anchor it and establish the slope which is important in calculating confidence limits.

The median effect has the greatest precision as an endpoint for quantal tests. Familiar examples are the *median lethal concentration (LC50)* and the *median effective concentration (EC50)*. Half of the test organisms would have shown the effect at concentrations lower than the endpoint (C). The other half of the organisms would only show effects at higher concentrations. The LC50s and EC50s always have a test duration associated with them (e.g., 96-h LC50)[13]. For an EC50, the particular effect being tabulated must also be stated.

In quantal experiments, the endpoint is estimated by fitting a line to the concentration-effect data (Figure 1). The LC50 could be read directly from an eye-fitted line (see Figure 1). More formally, the estimate is usually made by "probit analysis" or by some other specialized line-fitting procedure which uses all of the data generated by the test. Accordingly, a line such as the one shown in Figure 1 is often referred to as a "probit line". The mathematical model of the concentration-effect relationship also describes the associated error term and estimates the precision of the endpoint, customarily expressed as the 95% confidence limits of the LC50 (or EC50).

---

[13] The term EC50 applies to a lethal effect or any quantal sublethal effect (e.g., whether *Daphnia* did or did not continue to swim). It must not be forgotten that an EC50 is a concentration, and has the chemical concentration units of the toxicity test associated with it. Also, the operative noun in the name is "concentration" and it is not necessary to write "EC50 value" any more than one would write the redundancy "concentration value".

There is no mathematical impediment to calculating lethal or effective concentrations for any proportion of the population (e.g., LC25, EC10). However, the greatest precision is associated with the median effect, and as values become further from 50%, rapidly increasing numbers of test organisms would be needed to maintain a given degree of precision. Accordingly, the LC50 and EC50 are commonly used as the most precise quantal endpoints for describing and comparing the potency of different toxicants.

Figure 1 is plotted with a logarithmic or geometric scale of concentrations. Having adopted a logarithmic series of concentrations for carrying out the test (Section 2.2), the analysis does not abandon that scale, but continues using it, as is proper. Calculations of probit analysis use the logarithm of concentration, and in fact the logarithms are the "real" values. Arithmetic values are shown on Figure 1 for the convenience of most people, who can understand an arithmetic value much better than a logarithm. The same principle applies to an endpoint derived from a test. The EC50 is actually a logarithm; it is converted to an arithmetic value for ease of comprehension by most investigators. It follows that any manipulations of endpoints from a toxicity test, such as averaging or statistical testing of significant differences between endpoints, should be done using the logarithmic values, and not the "translated" arithmetic values (e.g., Section 3.2.2) [14].

Exposure time is an important component of all environmental toxicity tests, and quantal tests are no exception. If the results shown in Figure 1 were taken to be those of an acute lethality test, a series of tests with different exposure times would be expected to result in a series of different probit lines. Short exposures would require higher concentrations to encompass mortalities from 0–100%. Long exposures would be expected to require lower concentrations. However, in acute testing, there is often an exposure time at which the maximum acute effect has been achieved. In other words, prolonging the exposure would not increase the magnitude of acute effect. A time-series analysis of the EC50s for various exposure times can be very revealing. This often demonstrates a "threshold" concentration for acute effect, presumably a concentration which the test organisms can deal with for an indefinitely long exposure. Such an *incipient EC50*, or *time-independent EC50* is clearly a useful and meaningful endpoint. Time-series analysis was recommended by a workshop on statistical analysisof toxicity tests (OECD, 1998a), and is discussed in more detail in Section 5.7.3.

2.5.2 Estimating Inhibiting Concentration in Quantitative Tests

This type of test measures the effect of test concentrations on a graded sublethal variable, such as the weight that organisms attain, or the number of offspring that they produce.

The preferred endpoint is the *Inhibiting Concentration for a (specified) Percent Effect (ICp)*. It is the toxicant concentration that is estimated to cause a specified percent reduction in the biological function of interest, compared to the control [15]. A value is selected for "p", usually 25% in Environment Canada tests, and the symbol becomes IC25. (An effect of 25% reduction sounds large, but is of the same general order as the reduction in performance which is usually found at the no-observed-effect concentration, as described in Section 2.5.3.) For example, if the control organisms in a toxicity test produced 1000 eggs, the investigator would use the results to calculate the IC25, the concentration estimated to be associated with production of 750 eggs. As another example, in the algal growth test with *Selenastrum capricornutum*, the endpoint is the IC50 for "growth inhibition", i.e., 50% fewer algal cells than in the control. This estimate is really based on populations attained (i.e., numbers of algal cells existing in the test containers), and accordingly is estimated as an IC50.

---

[14] Failure to follow the principle of calculating with the logarithms of concentration is common, unfortunately. Many computer programs for calculating toxicity endpoints require entering arithmetic values and fail to convert them to logarithms, which biases the estimate of the EC50 or IC25. Investigators must enter concentrations into such a program as logarithms, for a correct estimate. Similarly, averages of several EC50s are often mistakenly calculated as the average of the arithmetic "translations", instead of the average of logarithms. To illustrate that such a procedure is not correct, the term *geometric mean* is sometimes used herein, which requires conversion back to logarithms of concentration, or else some equivalent mathematical procedure. Failure to use a logarithmic scale of concentrations is not limited to estimates of initial endpoints such as EC50. The mistake is sometimes seen in otherwise sophisticated models of data, intended to advance the field of environmental toxicology; such a basic error means that the general conclusions are questionable and should be re-examined (e.g., Baird *et al.*, 1996).

[15] This endpoint is not interchangeable with EC50. In quantitative tests suitable for measuring the ICp, each organism or group of organisms shows continuously variable performance in a function such as number of eggs produced. The percent reduction below control performance is the endpoint. An EC50 might be estimated from such a test by twisting the data into quantal form, such as whether a given individual did or did not produce eggs. Obviously, much information would be lost in a quantal treatment. The distinction between the two types of tests should be

ICps are being increasingly used to express toxicity in both short and long sublethal tests that have a continuously measurable biological variable such as growth or reproduction. The ICp is a point-estimate on a linear model of the test data, and so, like the EC50, it can assume any value, not just those concentrations selected for use in the test. (See Section 2.5.3 for discussion of the NOEC/LOEC which can *only* be concentrations that were used in the test.) Among other advantages is that an estimate of error can be calculated for the ICp. No systematic bias in the magnitude of estimated ICps will be caused by variable response, poor experimental design, or imprecise procedure; such faults will be reflected only in wider confidence limits.

Despite the advantages of the ICp, the current method of estimating it by non-parametric linear interpolation should be improved, especially for lower values of p such as IC10. The ICp is not estimated by fitting a line to the complete relationship between concentration and percent effect, but by interpolating between the two measured effects which lie, respectively, above and below the value "p". In that respect the estimate fails to make use of much of the information from the test, and the estimate is open to more potential variation. When better parametric or non-parametric methods are developed and proven, they should be adopted.

2.5.3 Estimating NOEC in Quantitative Tests

The ***null hypothesis approach*** is a totally different way of analyzing quantitative sublethal data. Quantitative measurements are made in the same fashion as for ICp, for example, the size attained by organisms at each concentration. The hypothesis is adopted, that the measured variable (size) in concentrations of the toxic material is not significantly different from the size measured in the control. Standard mathematical testing is carried out to test the hypothesis at a stated level of statistical significance, usually a probability of 5%. If the hypothesis is disproved, mathematical testing is continued to discern which concentrations yielded effects that were significantly different from the control, and which did not.

The *Lowest-Observed-Effect Concentration* (*LOEC*) is the lowest concentration in the tested series, at which a biological effect was observed (e.g., size significantly lower than the control). The *No-Observed-Effect Concentration (NOEC)* is the highest concentration at which the adverse effect was not observed (e.g., size same as the controls); it is always the next lowest concentration in the dilution series, after the LOEC.

The *Threshold-Observed-Effect Concentration (TOEC)* may be calculated as the geometric mean of the NOEC and LOEC, for the convenience of having a single value to describe toxicity, rather than two. As the name implies, the TOEC is intended to be an estimate of the threshold, or concentration at which toxic effects "just begin" [16].

---

maintained in the name used for the endpoint, so that a correct impression of the type of test is conveyed by a report. Some European and U.S. methods outline calculations for an IC50, then mistakenly call it an EC50, a practice that is prevalent in methods with algae and bacteria. Unfortunately, OECD makes this mistake in many of their otherwise authoritative documents on standard methods, thereby misleading new investigators.

[16] Preceding Environment Canada test methods call this geometric mean the TEC for "threshold-effect concentration". That was perhaps an unfortunate choice of wording since there is an established nomenclature in which TEC signifies the "true" threshold, as opposed to the *observed* or estimated threshold in a given experiment of a particular power. NEC, for example, is used in environmental toxicology to mean the "true" concentration which causes no effect, as distinguished from the NOEC seen in a test. It is recommended here, that the acronym should be changed to TOEC, signifying *threshold-observed-effect concentration*.

The term *Maximum Acceptable Toxicant Concentration*

There is a temptation to interpret the NOEC as if it meant there were no "real" effects at that concentration, but that is not exactly correct. There would likely be real or "true" effects at the NOEC, not detected by the statistical analysis, which is always limited by the power of the experiment. The magnitude of true effect that was not significant would therefore vary with the design and power of the experiment. To give an extreme illustration, in one experiment the NOEC might represent a 5% effect on the organisms, and in another test the NOEC might represent a 50% effect. Some actual average reductions in performance at the NOEC compared to the control, are listed below, having been estimated from analyses of data-sets for fish available in the literature (Suter *et al.*, 1987).

parental survival	22% reduction
fecundity	42% reduction
hatching success	12% reduction
larval survival	19% reduction
weight	20% reduction

There is a recent strong movement away from the NOEC/LOEC combination as an endpoint in sublethal tests [17]. The chief criticisms are three-fold. First, the possible values of NOEC and LOEC are limited to whatever concentrations happened to have been chosen by the investigator. In other words, the NOEC and LOEC are not point estimates like the ICp. Second, as discussed previously, the particular concentrations which emerge as LOEC and NOEC, from tests of statistical significance, are very much governed by the design and power of the experiment. A test using very few organisms, and having high variability which might have been caused by poor design or procedures, will estimate relatively high values of NOEC and LOEC compared to a low-variation test with many replicates and many organisms. Third, there is no way of estimating confidence limits about the endpoint, as can be done with point-estimation techniques. Valuable information in the dose-response curve is not used (e.g., the slope). In fact, NOEC and LOEC can often be calculated in the absence of a consistent concentration-effect curve, but the utility of the calculated values is questionable if so derived.

Accordingly, for those investigators following test methods published by Environment Canada, it is recommended that ICp should be used as the primary measurement of effect, when it is valid and suitable. The NOEC/LOEC technique could be used as a secondary measurement. Investigators should keep up with new developments as it is likely that improved methods will become available for estimating ICp, and that other endpoints based on linear modelling will also be developed. A new guidance document on statistical methods contains additional information on estimating endpoints (EC, in prep., a).

(MATC) has been used in past years as an empirical endpoint for sublethal life-cycle tests. Various authors have confusingly used the term to signify NOEC, the geometric mean of NOEC and LOEC, some undetermined value between NOEC and LOEC, or the range NOEC to LOEC. Recent literature tends to abandon MATC and instead, use the specific terms NOEC and LOEC. The term MATC might well disappear; its use of the word "acceptable" would be more suitable for a quality criterion that incorporated human judgement, rather than for a calculated endpoint of a test.

[17] There is general accord in publications on this topic. The publications note the deficiencies of the NOEC/LOEC technique, and sometimes it is recommended that any estimates be accompanied by statements of the minimum observable difference under the conditions of the test. Generally a linear model is recommended, such as ICp, with recognition of the deficiencies of present methods of analysis. Tentative suggestions are sometimes made for other analytical techniques, as yet undeveloped. The movement towards linear models instead of those based on null hypotheses started with Stephan and Rogers (1985) and Suter *et al.* (1987). Other papers are by Bruce and Versteeg (1992), Hoekstra and Van Ewijk (1993), and Suter (1997). A workshop conducted mainly by scientists from the Netherlands was clearly aimed at influencing standard methods of analyzing tests in Europe (Noppert *et al.*, 1994). The OECD itself commissioned a study by Pack (1993), who followed the pattern with a strong recommendation to move away from methods based on NOEC towards a linear model with point-estimates. The same recommendation came from an OECD (1998a) workshop on statistical analysis of toxicity data. A major workshop in the USA provides considerable discussion and lists similar reasons benefits of using a point estimate such as ICp or EC50 (Chapman *et al.*, 1996).

2.6 Precision of Tests and Variation Between Laboratories

Key Guidance

- *Accuracy cannot be measured for a toxicity test, although use of a reference toxicant can lend some credibility.*
- *Within-laboratory precision is often 20–40%. Interlaboratory precision can be expected to be reflected in coefficients of variation of 30–50%. That is similar to results in other fields of toxicology and chemical analysis.*

The need for accuracy and precision has provided part of the impetus for Environment Canada to publish standard test methods as have other organizations that set standards in the USA and Europe. Modern sublethal aquatic tests are satisfactory in their variation; the coefficient of variation often ranges between 20–40%, which is similar to the variation found for chemical measurements of toxicants in water.

Accuracy of a toxicity test cannot be measured. Accuracy signifies the closeness of the test result to the true value. Unlike chemical measurements, there is no independent method of measuring the true toxicity of a material, except by a second toxicity test, in which case the accuracy would not be known for either the first or second test.

It must be assumed, therefore, that each test is providing an accurate result for the particular conditions of the test (type of dilution water, robustness of the test organisms, size, season, etc.). Obviously there could be a spectrum of accurate answers depending on which modifying conditions are active. Modifying factors presumably account for much of the variation in toxicity results, both within and among laboratories. For toxicity tests, all the observed variation is customarily assigned to the category *precision*, not accuracy, or simply "variation".

The closest approach to measuring the accuracy of toxicity tests is to use a reference toxicant. The results can alert an investigator to apparent changes in the laboratory (Section 3.2.2). The accuracy still cannot be measured, though, as it remains unknown for the test with reference toxicant. It is also unknown if an active modifying factor (e.g., season) would influence the action of a reference toxicant (e.g., phenol) in a parallel fashion to its influence on the toxicant of interest (e.g., a metal-mining effluent).

Precision is a measure of the relative closeness to each other, of measurements repeated in the same manner. For toxicity tests, repeated measurements of the same toxicity tests within a given laboratory are usually taken as a measure of precision. A study with five sublethal marine tests (Morrison *et al.*, 1989) is typical in its conclusion that good biological testing has precision similar to that found in chemical measurements. About as good precision as might be expected for aquatic testing was published for 25 tests with copper and rainbow trout over two years, yielding a range of LC50s from 274–381 μg/L, a 1.4-fold range and a coefficient of variation of 9% (Dixon and Sprague, 1981a).

Interlaboratory variation is of great practical interest. Several major studies indicate that in any compilation of toxicity endpoints from different laboratories, for one material and one aquatic species, one should expect to find a five-fold range in estimates, and sometimes a ten-fold range (reviewed in Sprague, 1985). Part of this variation is explained by modifying factors such as the type of dilution water, but much less variation is certainly desirable. Chapman (1995) considers that an approximate twofold range in results between laboratories would be the acceptable limit.

In some recent interlaboratory studies of aquatic toxicity, the reported coefficients of variation (CVs) ranged from 13–<60%, but were mostly in the range of 30–50%. [18]

---

[18] Variation in 7-d tests with larval fathead minnows was assessed by ten U.S. laboratories (API, 1988). There was good agreement, and interlaboratory CVs (the standard deviation as a percentage of the mean) were 13% for results based on survival of larvae and 52% for results based on weight. A very large compilation for comparable tests with fathead minnows on 141 effluents, from different laboratories, showed that 82% of comparisons had CV ≤40%, while 75% of cases had CV ≤30% (Rue *et al.*, 1988). The same compilation showed that in within-laboratory testing of 46 effluents, almost all (89%) of the effluents had CV ≤30%. Another comparison of sublethal tests with fathead minnow larvae at nine laboratories in the San Francisco area reported a coefficient of variation of 31% (Anderson and Norberg-King, 1991).

For comparison, a coefficient of variation of 30% might correspond to approximately a four-fold range between the lowest and highest endpoints (e.g., if mean LC50 were 10 mg/L, the range might be 5–20 mg/L).

The variation in aquatic toxicity tests can be considered as similar to that for other kinds of toxicology, or for most chemical measurements. For example, in drug-testing with rats, Henig (1979) mentions a seven-fold variation in results, which could be caused by changing the method of housing the rats. For chemical analyses of 126 priority pollutants at the same concentrations as the USEPA criteria for fresh water, the between-laboratory CV was ≤70% for almost all (90%) of the measurements, and <40% for about half of the cases, similar to biological CVs (Rue *et al.*, 1988). The USEPA (1991a) also concludes that chemical and biological precision is entirely comparable, especially for organic chemicals and inorganics near the chemical detection limit.

Although the biological data may be similar to measurements in environmental chemistry, it would, of course be desirable to have much less variation among laboratories. In an overall environmental context, the smaller the variation among laboratories, the more meaningful the results. Comprehensive examinations of the causes of variation in aquatic tests and which variables might deserve attention in any effort to improve variation are provided by Burton *et al.* (1996) and Ausley (1996). Infection with pathogenic micro-organisms is an obvious possible cause, and has been documented in tests with fathead minnows (Kszos *et al.*, 1997).

A couple of implications arise from these evaluations. (1) Major portions of the variation do not have obvious explanations at present, and so toxicity endpoints must not be interpreted as absolute values. (2) If it is desired to derive an environmental quality guideline or regulatory limit from a list of determinations of EC50s or IC25s for a given species, it would be quite unreasonable to pick the lowest value and use it as an estimate of toxicity. It might be an unusual outlier to the same extent as the highest value in the list. A geometric mean value for the species would usually be a better estimate, and such averaging is now standard in deriving water quality guidelines (Stephan *et al.*, 1985).

Confidence limits from an individual toxicity test must not be mistaken for any sort of overall range for the LC50 of a material. They are only internal estimates of variation for a particular test with a given group of organisms, under the conditions at that time.

For 7-d tests with a Pacific mysid crustacean, the CV of intralaboratory LC50s was 7%, and was 19% for growth Hunt *et al.* (1997). Between laboratories, the CV for LC50s was 14% and 15% for growth, all relatively low values (i.e., good precision). Other assessments of toxicity test variability have reached similar conclusions (Buikema*et al.*, 1982; Nebeker, 1982; Qureshi *et al.*, 1982; Grothe and Kimerle, 1985; Pickering, 1988; USEPA, 1991a).

Five Canadian laboratories did studies of Environment Canada's 14-d static toxicity tests with the amphipod *Hyalella azteca* (EC, 1997b), on four samples of field-collected sediment. At the various laboratories, the CVs for mean survival ranged from 3.6–19.6%, and from 28.4–48.8% for growth. A similar interlaboratory comparison was made using the 10-d test of survival and growth of the midge *Chironomus riparius* (EC, 1997a). The CVs for survival ranged from 2.1–25.1%, and from 7.7–27.4% for growth (Milani *et al.*, 1996).

Similar studies were done at 9 to 11 U.S. laboratories using U.S. sediment tests (Burton *et al.*, 1996). For mortality, the overall CV was 39% for *Hyalella*, and 14% for *Chironomus tentans*; for growth of the midge, CV was 32%. Tests with the reference toxicant, using water without sediment, showed CVs of 16 and 20%. The authors concluded that these test methods had "relatively low variance and acceptable levels of precision in interlaboratory comparisons".

Test Methods and Quality Assurance

Key Guidance

- *Quality assurance (QA) is a system of management and operational activities designed to ensure adequate control of quality in the work produced by a laboratory.*
- *Quality control is part of QA, involving routine checks and calibrations in normal operations.*
- *Programs of QA are becoming more formal, with Canadian accreditation programs being developed for testing laboratories.*

Techniques for assessing the reliability of a test result are described in this section. Reliability of data is synonymous with their quality, which is indicated by the practices of the test laboratory in *quality assurance* (QA) and *quality control* (QC). Some components such as the health of the organisms could have a very direct effect on the outcome of a test, whereas general operating procedures of the laboratory might have an indirect influence on results.

A QA program assures that steps to maintain and improve data quality are in place and that the limits of uncertainty associated with the data are known (Sergy, 1987). *Quality assurance has been defined as an overall system of management and operational activities designed to ensure adequate control of quality.* The objective of QA is to provide data that are complete, precise, representative, and comparable. ***Quality control is a component of QA, and is a system of routine checks and procedures included in normal operations*** (such as toxicity tests) to ensure that data of good quality are produced within a laboratory.

Quality assurance and control practices are inherent in all toxicity tests. Rigorous application of the practices is a major factor distinguishing a *standard* biological test. Information on QA makes it possible to judge whether a result is, indeed, derived from such a standard test. Quality assurance retains its importance in identifying tests that do not satisfy some procedural criterion, or tests in which the results are unusual or might be questioned. If a test is unique (no opportunity to repeat it), QA information can be used to evaluate its reliability.

Requirements for QA in sample collection are not discussed here but they are covered for aqueous samples in each Environment Canada method document, and for sediment samples by EC (1994c) and Mudroch and Azcue (1995). Guidelines for statistical analysis will be addressed in *Guidance Document on Statistical Methods for the Determination of Toxicity Test Endpoints* (EC, in prep., a).

3.1 Quality Assurance in Toxicity Testing

Most users of toxicity testing services see only the directly pertinent information in a report generated by the test facility. Many other aspects of laboratory operations, collectively termed QA, can influence the reported result. Such background information should be accessible on request, to aid in interpreting the results.

A QA program for toxicity tests will include day-to-day operation of the laboratory, source and condition of the test organisms, implementation of methods, training of personnel, sample collection, handling, disposal, co-ordination of testing, procedures for data collection, analysis, and reporting. *Written standard operating procedures for a given laboratory are a main component of all QA programs.*

A second component of a QA plan is QC practices. Quality control embraces procedures to measure quality of data and includes standardization, calibration, replication, controls, reference toxicants, split samples, participation in interlaboratory testing programs, and similar activities. Quality control provides direct quantitative measurements of data quality, and also monitors the proper functioning of the test system.

Site inspections (or audits) can assess the adequacy of a QA program, and are usually a key component of accreditation programs. An audit is a formal verification of the ability of a laboratory to produce valid results. It covers all the facilities, procedures, documentation, and ability of personnel mentioned in this section, reviews the quality of systems in the laboratory, and requires the laboratory to meet specific levels of performance. In a formal audit, the test facility first submits information to the auditors [e.g., quality control manual or standard operating procedure (SOP)]. Next, a site visit is usual, to verify the information. The facility is then rated according to predefined criteria of performance. Deficiencies are identified and recommended or required improvements are listed. Guidance documents are available for auditing biological laboratories, complete with checklists for specific tests (Klemm *et al.*, 1990; EC, 1993a). If a user of test results wished to audit a potential "supplier" laboratory, a strong component should be a site visit, formal or informal.

Certification and accreditation provide two avenues for recognition of a laboratory's qualifications. Certification is a recognition that the facility can conduct a particular test, usually by means of a written submission and/or performance in testing samples. Accreditation is a more rigorous evaluation and often includes tests of performance as well as a site visit. Even for results received from a certified and accredited laboratory, the user would be well advised to assess the quality of data; regulatory agencies customarily hold the user responsible for ensuring that information is valid.

The *Canadian Association for Environmental Analytical Laboratories* (CAEAL) is a third-party, non-governmental accreditation body. It has a program for accrediting toxicological laboratories, the first of its kind in Canada. Accreditation requires participation in tests of performance, and an audit of the laboratory every two years. Since 1994, accreditation has been available for three toxicity tests published by Environment Canada: acute lethality for rainbow trout and *Daphnia magna*, and bacterial luminescence in the liquid-phase test. By 1998, CAEAL offered accreditation for 14 toxicological tests.

3.1.1 Related International Standards of Laboratory Practice

Good Laboratory Practice (GLP) is a program of QA originally developed by the Organization for Economic Co-operation and Development (OECD). It is intended for testing new products for registration, particularly commercial chemicals and pesticides. Guidelines for GLP have been endorsed by OECD member nations, but not all of the countries have enacted GLP regulations, as yet. The guidelines describe the requirements for conducting and reporting specific studies. The level of QA in GLP studies will usually meet requirements for ensuring reliable data. The Canadian Environmental Protection Act requires that toxicity testing for registration of new substances must conform with OECD standards of GLP (EC and HWC, 1993) [19].

Under the auspices of the International Organization for Standardization (ISO), a series of QA standards, collectively referred to as "ISO 9000", is becoming more widely used by Canadian manufacturing and service industries. The standards might become increasingly relevant to toxicity laboratories. They cover methods used to design, produce, install, inspect, package, and market products and services. There are five categories of standards.

9000	Guidelines for selection and use
9001	QA in design, development, production, installation and servicing
9002	QA in production and installation
9003	QA in final inspection and test
9004	Quality management and quality system elements

Secondary suppliers must be ISO-certified in order to supply a company certified under ISO 9000. For "supply" laboratories, the required technical standards are outlined in ISO/IEC Guide 25 ("General requirements for the competence of calibration and testing laboratories"). Laboratories that meet the requirements of Guide 25 comply with the requirements of ISO 9002 when they supply test results.

---

[19] EC and HWC (1993) gives sources of OECD publications. The OECD has eight broad-ranging publications on components of GLP (OECD, 1992e-i; 1993b, c; 1994). In the USA, the EPA has the Federal Insecticide, Fungicide, and Rodenticide Act (USEPA, 1989d) and the Toxic Substances Control Act (USEPA, 1989e). The former Act includes toxicological and health-related studies, mostly mammalian, and the latter covers both health and environmental studies. The GLP programs are expanded from the first GLP regulations issued by the U.S. Food and Drug Administration in 1978.

The CAEAL has been following ISO Guide 25 during site assessments, under the authority of the Standards Council of Canada, which is the accrediting body in Canada for ISO. Recently, the Canadian Standards Association has published a standard for accrediting Canadian laboratories (CAN/CSA-Z753, "Requirements for the competence of environmental analytical laboratories"). This has been adopted by the Standards Council of Canada and the CAEAL. Standard CAN/CSA-Z753 combines the requirements of ISO 9000, ISO/IEC Guide 25, and other national and international quality standards, for laboratories testing environmental samples and calibrating equipment for environmental monitoring. It has been designed for internal use of environmental analytical laboratories, as well as for agencies granting accreditation.

3.2 Test Methods

Key Guidance

- *A methods document should encompass the following topics. (1) Introduction with rationale and applications. (2) Test organism, sources, culture or holding, and health. (3) General design and replication, details of exposure, test vessels, volumes and densities. (4) Method modifications for special types of samples. (5) Preparation of sample. (6) Test observations and measurements. (7) Analysis of results. (8) Response in controls, reference toxicant, and other criteria of acceptability. (9) Requirements for reporting.*
- *If it were desired to use an alternative method, its adequacy could be rated by examining the procedure for each of the topics listed above.*
- *Results of tests with a reference toxicant should be followed by plotting sequentially on a warning chart. Warning limits should be calculated as two standard deviations from the mean, using logarithms of concentration for all calculations.*

This section enumerates the components of standard test methods, exemplified by Environment Canada methods. It focuses on reference toxicants as a method of quality assurance and it ends with a framework for comparing different methods.

There are hundreds of methods for testing toxicity but only a few are fully standardized. Standard methods thoroughly document all procedural steps and facilities, and have a number of common desirable features (Sections 3.2.1 and 3.2.3). Methods have been developed internationally by the ISO, the OECD, and in North America through the American Society for Testing and Materials (ASTM). The Environment Canada method documents for aqueous and sediment samples follow a standard pattern, and are described herein as an example of good coverage for a test. The documents are developed in consultation with numerous toxicological experts and practitioners within and beyond Canada.

3.2.1 The Standard Pattern of Environment Canada's Test Procedures

The table of contents in each test method published by Environment Canada provides an overview of what to look for in a test method; it can be used as a checklist to evaluate other test methods. The following six topics are addressed in order.

- Introduction
- Test Organism
- Test System and Facilities
- Universal Procedures
- Specific Procedures
- Reporting Requirements

An introductory glossary defines relevant terminology. The introduction provides rationale, reviews previous similar methods, and provides general information on the test species including taxonomy, distribution, and ecological relevance. The applications and limitations of the test are specified. Environment Canada offices that can clarify interpretations are listed by address in an Appendix.

The next section (Test Organism) is devoted to the test species. It gives potential sources of the organisms, and holding conditions including acclimation, feeding, temperature, lighting, photoperiod, and water quality. Tables summarize this information, and there are checklists of holding conditions for purposes of quality assurance.

Criteria for health of organisms are specified. For example, mortality of rainbow trout must be <2% during the week before a test. Considerably more complex criteria including reproduction apply for some tests (e.g., *Ceriodaphnia*). These criteria are intended to identify relatively unstressed organisms that should give a valid result.

The Test System and Facilities section then defines the basic requirements for the test system. The designs are intended to reduce variability and minimize stress other than from the toxicant. Physicochemical conditions during the test are usually similar to those for culturing or holding the organisms. Items specified include location within a laboratory, amount of space, type of vessel, lighting (type and photoperiod), temperature, and quality of dilution water.

The universal test procedure is then described in each document (Universal Procedures). Distinction is made between absolute requirements (that *must* be done for a valid test) and preferred procedures (that *should* be done). Initial setup involves selection of a diluent, control, and treatment series; the actual test is then described step by step. Specified test conditions include those in the preceding paragraph, and condition of organisms, feeding if any, physicochemical measurements (e.g., temperature, dissolved gas and aeration rate, pH, specific conductance, ammonia), observations of effect, duration, endpoints, criteria for validity, and reference toxicants (see Section 3.2.2). Conditions are summarized in tables that serve as checklists for assessing correct conduct of the test. Methods are specified for calculating the endpoints. More detailed guidance on statistical methods is to be covered in a separate document (EC, in prep., a).

For aquatic tests, separate sections give applications and modifications of the universal test procedure, usually for testing of chemicals, effluents, elutriates, leachates, groundwaters, and receiving waters. Any additional information required on handling and testing these kinds of samples is provided along with any modifications of criteria for a valid test.

The same general format is followed for tests with sediment. There is specific guidance on how to obtain and handle samples, generally a more complex and demanding operation than with water samples. There are two major groupings of specific methods, those for testing a field-collected sediment or a similar solid substance, and those for testing chemicals introduced into sediment.

The final section of each test method (Reporting Requirements) outlines (a) the minimum requirements for a given report, and (b) additional information which must be included in a general report or held on file for a minimum of five years (see Section 3.3.1). Subsequent appendices give supporting or background details. Usually this includes instructive information on the species and its environmental requirements, on the details of preceding test methods elsewhere, and on the procedures for analyzing results.

3.2.2 Reference Toxicants

Every toxicity method published by Environment Canada recommends use of one or more reference toxicants (Table 2) in the last part of the Universal Procedures section. Reference toxicants are discussed here as an important part of a QA program.

A reference toxicant is a standard chemical used to measure the precision of results for a given test method within a laboratory. It is a tool for determining whether a laboratory can obtain consistent, precise results, providing a check on functioning of the test system. A reference toxicant can also measure sensitivity of organisms, as a base for results with the test material. Canadian guidance documents for reference toxicants are available for aquatic tests (EC, 1990d), or for sediments and other solids (EC, 1995a).

For many aquatic tests, Environment Canada specifies zinc as an inorganic reference toxicant, and phenol as an organic substance (Table 2). These two toxicants are supplemented or replaced by other substances for some species, media (sediments and soils), or endpoints (sublethal) as shown in Table 2. Eight primary criteria were used in selecting appropriate reference toxicants: previous use, availability in a pure form, solubility, stability in solution, stability during storage, ease of analysis, stable toxicity with normal changes in qualities of laboratory water, and ability to detect abnormal test organisms (EC, 1990d). There were not enough data to make good use of the last criterion. The other criteria could be used to evaluate additional potential reference chemicals.

Table 2 **Summary of Reference Toxicants used in Environment Canada Test Methods**

Test Organism	Method Number	Reference Toxicant		
		Zinc	Phenol	Other*
Bacteria				
Photobacterium phosphoreum (now *Vibrio fisheri*)	1/RM/24	yes	yes	$K_2Cr_2O_7$
Plants				
Selenastrum capricornutum	1/RM/25	yes	yes	NaCl
Lemna minor	1/RM/37	no	no	$K_2Cr_2O_7$, K_2CrO_4, KCl
Invertebrates				
Daphnia spp.	1/RM/11	yes	no	NaCl, $K_2Cr_2O_7$
Ceriodaphnia dubia	1/RM/21	yes	yes	NaCl
Sea urchins/sand dollars	1/RM/27	no	no	$CuSO_4$, $CuCl_2$
Marine amphipods	1/RM/26	no	no	Cu, Cd, $C_{16}H_{10}$
Hyalella azteca	1/RM/33	no	no	$CuSO_4$, $CdCl_2$, NaCl, KCl
Chironomus spp.	1/RM/32	no	no	$CuSO_4$, $CdCl_2$, NaCl, KCl
Vertebrates				
Rainbow trout	1/RM/9	yes	yes	no
Threespine stickleback	1/RM/10	yes	yes	no
Fathead minnow	1/RM/22	yes	yes	NaCl
Salmonid fish	1/RM/28	yes	yes	no

* Cd – cadmium, Cu – copper, $CuSO_4$ – copper sulphate, $CdCl_2$ – cadmium chloride, $C_{16}H_{10}$– fluoranthene, KCl – potassium chloride, K_2CrO_4 – potassium chromate, $K_2Cr_2O_7$ – potassium dichromate, NaCl – sodium chloride

Reference tests should be frequent enough to detect any change within a culture or batch of organisms, or between batches. A warning chart is used to track any changes (Figure 2). For organisms that are cultured in-house, monthly tests are usually adequate. Organisms received in batches should be tested when first received, during the holding period, and just before they are all used for toxicity testing. If a small group of organisms is collected in the field or purchased, for use in a few tests, then reference toxicants should be run with each test; that should also be done for non-routine types of tests. Some monitoring or regulatory programs might have specific requirements for frequency of reference tests.

Current data for the reference toxicant are plotted along the time axis of a warning chart (Figure 2). Also plotted are the historical geometric mean, and *warning limits* of the 95% confidence limits, calculated as two standard deviations from the mean, using logarithms of concentration for the calculation [20]. New results are then compared with the limits as shown in Figure 2. The calculations and plotting on

---

[20] Test concentrations should be in a geometric series (Section 2.2). All analyses of results should continue to use logarithms of concentration; the true values of EC50s and IC25s are logarithms but the endpoints are customarily converted to arithmetic values because most people comprehend them much better (see Section 2.5.1). Accordingly, in the control chart for a reference toxicant, the mean LC50 is obtained as the average of logarithms, and the standard deviations are similarly derived from a set of logarithmic LC50s.

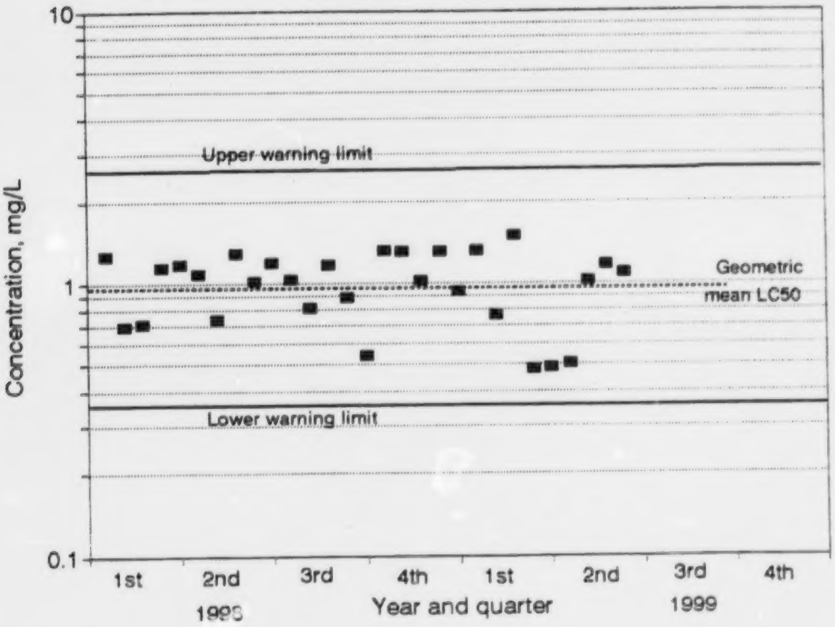

Figure 2 **Warning Chart for Tests with a Reference Toxicant.** The warning limits are two standard deviations from the historic mean of the median lethal concentrations (all calculated using the logarithm of concentration). This figure represents a relatively regular pattern for a testing laboratory, and all LC50s are well within the warning limits.

semi-logarithmic paper could be done by hand, or a spreadsheet program could be set up for calculations, with sequential replotting of the endpoints and limits by the computer.

On average, one in twenty endpoints will fall outside the 95% limits by chance. For repeated tests with the same dilution water within one facility, it should be possible to achieve a 30% coefficient of variation [21], considered acceptable (EC, 1990d). If results for a reference toxicant fall outside the warning limit, the laboratory should investigate and attempt to determine the cause(s). They might be benign, such as incorrect calculation of the result, or they might be serious, such as a stressed population of organisms which yielded unusual results.

The magnitude (span) of the warning limits is obviously important. Narrower limits would signify less variation in results obtained by a laboratory. Accordingly, the precision over time could be judged from a laboratory's warning chart, but no particular magnitude for the limits appears to have been designated as satisfactory for toxicological work.

3.2.3 Evaluating Other Methods

There can be a choice of methods for a given species. Methods for a 96-h lethality test with rainbow trout have been published by Environment Canada, British Columbia, Alberta, Ontario, USEPA, ASTM, APHA *et al.*, and the OECD. Sometimes there might be reason to use an alternative method, or to use the same test for two purposes (e.g., a test conducted according to a method published by Environment Canada might be used to satisfy a provincial regulation). If an alternate method is used, a practical comparison could use both methods on a proportion of the samples, and/or a reference toxicant. A relatively simple "paper" comparison can assess the similarities and differences of methods.

---

[21] The coefficient of variation is the standard deviation divided by the mean, expressed as a percentage. Sometimes ±3 standard deviations are also plotted on the warning chart as 99% confidence limits, and called *control limits*. A reference LC50 outside the control limit should trigger an immediate investigation, since that result would be expected by chance in only 1 of 100 tests.

A published and readily available method is the first essential component for any toxicity test. The method should provide enough detail to allow someone trained in toxicity testing to repeat the test, analyze the results, and determine whether the result is valid.

If the written method is satisfactory, the next step is to list the procedural items for ready comparison. Ideally, toxicity test methods, other than those of Environment Canada, should contain the following basic elements.

1. an introduction outlining the rationale for the test and a concise statement of its limitations and applications
2. information on the species of test organism, sources, culture or holding conditions, and criteria of health
3. details of test design, including exposure conditions, test vessels, loading densities, and replication
4. instructions for preparing the sample
5. frequency and type of observations during the test
6. procedures for analyzing the results
7. limits for response of controls and other criteria for acceptability of test

The above-listed items might also be used to evaluate methods with different test species, although two species cannot be expected to have the same relative sensitivities to a spectrum of toxicants. If the comparison of methods had important implications, a qualified environmental toxicologist should make the review.

Environment Canada (EC, 1994d) proposed a *specific and structured method* of evaluating environmental toxicity test methods for assessing contaminated sites. Each test was first ranked according to three criteria considered absolutely essential:

1. a readily available printed method;
2. specified reference toxicant; and
3. criteria for acceptability of a result.

Tests that met the essential criteria were assessed for 12 secondary but desirable attributes. These were:

1. organism identified to species;
2. measurable endpoint;
3. defined attributes of organism (age, weight, size);
4. number of organisms and replicates;
5. the frequency of observations;
6. volume of test solution;
7. volume of vessel;
8. collection, storage, and preparation of test solutions;
9. culture and handling of organisms;
10. environmental conditions during test;
11. diluents and nutrient solutions; and
12. methods for statistical analysis.

Weighted scores were then assigned (EC, 1994d). A test was considered complete if it met all the essential criteria and scored >88% on the desirable items. Other information was reviewed, including trophic level, reproducibility, field validation, and relevance to the Canadian environment. If a test met the essential criteria but scored <88%, it was considered a prototype; that category also included tests failing an essential criterion but scoring above 88%. The other tests which failed essential criteria and scored <88% were classified as being under development.

3.3 Test Reports

Key Guidance

- *A report on a toxicity test should contain the following information. (1) Type of sample and information on collection, shipping, and storage. (2) Description of test method, design, and conditions. (3) Any deviations from the standard test method. (4) Observations during the test including controls. (5) Graphical presentation of results. (6) Endpoints and the statistical procedures used to calculate them. (7) Information on the test laboratory, people, and time of test. (8) Data for a reference toxicant.*
- *Evaluating the reliability of a test from its report is largely a matter of checking for complete and satisfactory coverage of each of these items.*

A test report must provide the results in a clear and complete fashion. The report must, however, also describe the test design and conditions, so that a user can determine whether the test was done correctly and whether the results are reliable.

3.3.1 The Environment Canada Pattern for Good Test Reports

The final section of each test method published by Environment Canada provides a good framework for reporting data. The following information is usually the required minimum.

1. type of sample, collection site, time, and person sampling, and additional pertinent observations (e.g., colour, odour, consistency)
2. procedure for shipping the sample, when received and by whom, its condition, and subsequent storage and preparation
3. brief description of methods or reference to the published document
4. basic design and conditions (temperature, lighting, photoperiod, aeration, etc.) and information on diluent/control water
5. any deviations from the standard test method
6. observations and measurements during the test
7. graphical presentation of the results
8. relevant endpoints and the statistical procedures
9. name and location of test facility, time of the test, and names of the people who conducted it
10. latest data for a reference toxicant and the warning chart

Additional information is generally available at the laboratory. Most facilities use a bench sheet to record measurements during a test, and if necessary, the first step in verification is to confirm that all the data have been transcribed correctly from the sheets to the report. The general QA procedures would not be included in routine reports (e.g., information on stock cultures, calibration of equipment, training of personnel). If necessary, these would be accessible in an audit of the facility. Environment Canada programs require this information to be held on file for at least five years.

3.3.2 Using a Report to Evaluate the Reliability of a Test

The following procedure is recommended for reviewing a report and assessing the validity of results. The assessment requires that the test be based on an adequate written method.

1. Ensure that the report is complete.
2. Check that requirements for validity have been met.
3. Check for deviations from specified conditions and procedures, and for an explanation by the laboratory.
4. Compare the calculated endpoints to original observations.
5. Check supporting information on quality assurance.

The first step is to ensure that the report is complete (see list in Section 3.3.1).

The second step is to check that the predefined criteria for a valid test have been satisfied, particularly the performance in controls. For an unusual case of non-validity, a comprehensive review of data is required to determine if results are still useful.

The next step is to check for deviations in the methods; the report should describe any deviations and provide an explanation. The items of procedure mentioned in preceding sections should be checked, and those items are specifically listed in each Environment Canada procedure. Deviations from "must" requirements will likely make the result invalid; deviations from "should" procedures probably will not but judgement must be used. If any conditions or measurements are missing or peculiar, the laboratory should be contacted for clarification. Only if the recipient of the test data is satisfied that there was no undue influence on the test should the result be accepted. [22]

The calculated endpoints should then be compared with the data. The best approach is to approximate the endpoints by plotting or making approximate calculations, then compare the reported values. If the agreement is not reasonable, the cause of divergence must be sought. If uncertainty remains, the test laboratory should be contacted; errors can occur in transcription of data.

The final step is to review the remaining QA information that accompanied the report. This should include information on test conditions and organisms. If recent reference tests fall outside the warning limits, this must have been noted in the report, and laboratory personnel should have started searching for sources of variation.

---

[22] Problems that invalidate test results occur occasionally at most laboratories. If there were a legal or regulatory requirement to provide results at specified times, with appreciable penalties for missing a report, it would be possible to split samples between two laboratories, to reduce the risk of a data-gap.

Interpretation of Toxicological Data

An estimate of toxicity from a laboratory test is tied to species and particular test conditions. Obviously, these are not direct measurements of effects in the environment, but rather, they can predict the potential detrimental effect of a toxicant. This section addresses questions on interpreting toxicity results, to make predictions to the field. Sections 4.1 to 4.9 deal with practices that apply to most toxicity tests and Sections 4.10 to 4.13 deal specifically with wastewater, sediments, soils, and individual chemicals.

In the laboratory, toxicity is normally measured under carefully standardized conditions to ensure precise results, as for other measurements of the property of a substance, such as volatility or biodegradability. Good scientific practice prescribes standardized laboratory measurements under controlled conditions of such variables as temperature, light, and oxygen. The relationships can then be used to predict events in nature where conditions fluctuate. The process is similar to other technical predictions, e.g., a series of standardized measurements of evaporation rates can predict real-world rates under changing temperatures and partial pressures.

How well do the laboratory tests predict effects in nature? Section 4.1 reviews some of the research addressing this question.

4.1 Field Validation

Key Guidance

- *There have been many attempts to validate single-species laboratory tests by comparing them with effects in biotic communities. Most of these were general comparisons, and a few were rigorous scientific field validations.*

- *One excellent validation found 88% agreement of toxicity tests on 43 effluents with effects/no effects on natural communities. Eight deliberate studies by the USEPA were criticized for methods of analysis, but re-analysis showed general agreement. Most other comparisons with field effects also showed agreement.*

- *Most controlled studies have shown agreement between toxicity results and effects on communities in artificial streams or other mesocosms.*

- *Literature reviews have also found general agreement between single-species tests on effluents, and community-level effects. Two reviews each found 84% agreement, based on 13 and 63 useful experimental comparisons.*

There has been sufficient research carried out to show that single-species toxicity tests are usually predictive of effects on biotic communities. Such comparative research has to overcome certain difficulties. Some easily foreseeable complications in a real waterbody would include the presence of more than one active pollutant, and fluctuations in severity. A greater risk from an ecological point of view, is using knowledge gained at one level of biological integration (i.e., the species level in a toxicity test), to predict what might happen at a higher or broader level of integration such as populations or communities. For example, reproductive success of individual waterfleas in the laboratory would not necessarily predict what would happen to populations of crustaceans in the wild.

Scientific studies to assess the predictive value of single-species laboratory tests by means of a simultaneous comparative study of effects in a functioning community are often termed ***validation*** or ***field validation***, without implying that there will be positive confirmation. Most validations have provided general correlations, and show a strong balance of agreement between the effects in toxicity tests and effects in communities.

The validation studies available in the literature differ in scientific rigour. Some important features in designing or assessing validation studies are summarized in Section 4.1.4.

4.1.1 Validations in Natural Waterbodies

One excellent experiment involved treating a small river in Ohio to a constant concentration of copper for two years, and assessing wild and confined fish and other biota (Geckler *et al.*, 1976). Life-cycle tests with fathead minnows in a streamside laboratory yielded good agreement with population studies of invertebrates, and with performance of fish held in cages in the stream, i.e., favourable validation. An unpredicted major surprise, however, was that unconfined wild fish showed strong downstream avoidance reactions, at concentrations lower than those causing physiological effects. Accordingly, this experiment showed that single-species toxicity tests agreed with, and were predictive of, the physiological effects that prevailed in the stream at given concentrations of copper. The laboratory tests did not predict the total ecosystem effect because of the avoidance reactions.

Several lessons can be learned from the experiment in Ohio. First, the predictive worth of the laboratory tests was greatly improved by using upstream water to prepare the test concentrations. Indeed, laboratory tests with a standard supply of water gave poor predictions because the water lacked the metal-binding substances that were in the river water. Second, there are good reasons for periodic field surveys in the receiving water, to check that predictions from monitoring the effluent are correct. Third, laboratory tests of avoidance or preference responses to copper could have been undertaken, and might have predicted the avoidance reactions of fish in the stream. That approach gave useful results elsewhere (Sprague *et al.*, 1965). It has been recommended that avoidance/preference tests should always be included as a discrete part of a toxicity investigation, because behavioural and physiological responses can be quite different (Sprague, 1981).

An excellent Canadian example of field evaluation was done for a New Brunswick base metal mine which had acid-forming ore and drainage. The results of four sublethal toxicity tests were compared with effects on benthic invertebrates, and fish populations and health. There was a strong statistical correlation between each toxicity test (*Ceriodaphnia*, *Selenastrum*, *Lemna*, fathead minnow) and effects measured in the field (AETE, 1998a).

Eight deliberate validation studies by the USEPA in the 1980s showed reasonably good agreement. The objective was to test whether 7-d sublethal tests with larval fathead minnows and *Ceriodaphnia* correlated well with observed conditions in polluted waterbodies in various parts of the USA. Usually, toxicity tests were done on samples of effluent and receiving water ("ambient tests"). Standard field surveys of water chemistry, resident fish, macroinvertebrates, and usually periphyton and/or plankton were done simultaneously, with hydrographic work to identify plume concentrations. The waterbodies included:

- a small creek in an agricultural area;
- a stream in agricultural Oklahoma affected by an oil refinery, a fertilizer plant, and a municipal treatment plant;
- a river in Connecticut and two in West Virginia receiving numerous discharges;
- a river in Ohio receiving municipal and refinery effluent;
- a river in Alabama which received waste from a municipality and three coke plants; and
- an estuary near Baltimore (Mount *et al.*, 1984; 1985; 1986a,b,c; Mount and Norberg-King, 1985; 1986; Norberg-King and Mount, 1986).

The EPA's methods of comparison and analysis have been severely criticized. However, re-evaluation with canonical correlation analysis showed reasonable correlation between results from sublethal toxicity tests and effects in field surveys. Apparently, the same conclusions could be drawn most of the time, whether the approach was toxicity testing or biotic surveys. Marcus and McDonald (1992) criticized, then re-analyzed the research, and wrote that the sublethal "*Ceriodaphnia* and 7-d fathead minnow tests used to determine ambient toxicity can provide useful information about biological community structure in stream waters where ambient toxicity has a controlling influence". Another re-analysis found that the most useful variables were neonate production by *Ceriodaphnia*, dry weight of fathead minnow larvae in the toxicity tests, and numbers of species of fish and invertebrates in the field surveys (Dickson *et al.*, 1992). The analyses "demonstrated that statistically significant relationships between ambient toxicity and instream impact existed", and that "the relationship between ambient toxicity and instream biological response is strong".

A high-quality study in a Kentucky stream yielded excellent correlation between results of 8-d tests with fathead minnow larvae exposed to effluent, and field work on the affected stream (Birge *et al.*, 1989). The larval tests indicated negligible effect at 36% concentration, and field work agreed, with no significant effect on fish or invertebrates at a stream concentration of 33%. A station with 53% concentration of effluent did not show an effect on fish, but had only 22 species of invertebrates compared to control values of 30 and 34, and a diversity index of only 2.5 compared to the clean-water indices of 3.3 and 3.7. Other toxicity tests in the river ("ambient") showed excellent correlations with surveys of macroinvertebrates (0.93 with diversity index and 0.96 with number of species).

In North Carolina, Eagleson *et al.* (1990) assessed 43 effluents and their receiving waters in a very objective manner. Toxicity of each effluent was measured by a 7-d test with *Ceriodaphnia*, then effects were *predicted* if the calculated TOEC would be exceeded at low flow in the receiving water. Observed effects in the waterbody were judged by statistically significant difference in the distribution of taxa of macroinvertebrates at the control and affected stations. The overall agreement of the predictions with the field observations were 88%, distributed as in Figure 3.

Researchers in Virginia found a threshold of 1.7% effluent for reproductive effects on *Ceriodaphnia*, agreeing with a threshold of effect of 2.0% for field surveys of macroinvertebrates (Pontasch *et al.*, 1989).

A study of metals from mining in Colorado showed disagreement. There were reduced numbers of salmonid fish in a stream at concentrations below the TOEC that were expected to have no effect, according to pre-existing data from sublethal tests in the laboratory (Davies and Woodling, 1980). *In-situ* tests with fish agreed with the field findings.

There do not appear to be any studies of ecological effects in a marine system, compared to the toxicity of an effluent discharged into that system (Schimmel and Thursby, 1996). However there is some comparison with *in-situ* toxicity tests of the sediment. Marine benthic invertebrate communities were compared with laboratory toxicity tests on benthic sediments for 43 sites in a bay receiving several effluents (major sources were a copper smelter, pulp mill, chemical plant, and storm drain). Toxicity tests measured bacterial luminescence inhibition, abnormalities in oyster larvae and amphipod survival, and showed good correspondence with moderately or severely altered benthic invertebrate communities (>60% with all three tests). Both measurements corresponded with levels of known contaminants in the sediments. Effects declined in the three toxicity tests with distance from the major sources of contamination, while the abundance of major benthic taxa increased along the same spatial gradient (Becker *et al.*, 1990).

Thus there is a preponderance of agreement between the community studies in the field and the single-species tests in the laboratory. There is an appreciable fraction of the studies showing disagreement (e.g., Figure 3), but the discrepancy is not great enough to warrant the criticism of Power and McCarty (1997) about the "myth of laboratory-to-field evaluation". The main weakness of the attempts at validations so far, do not seem to be in any inherent lack of correlation, but in the methods adopted by investigators, in using general or subjective comparisons rather than rigorous tests of predictive ability (see Section 4.1.4).

4.1.2 Validations in Mesocosms

Some validation has used communities in outdoor artificial streams. At a USEPA research station in Minnesota, there are eight artificial channels, each 520-m long, allowing replicates. Natural water from the adjacent Mississippi river flows through the channels. A good example of validation in the outdoor streams is provided by 12-week exposures to pentachlorophenol (Zischke *et al.*, 1983). Laboratory tests estimated that 48 μg/L would be protective as a 30-d average concentration, (Stephan *et al.*, 1985). That was in reasonable agreement with the stream studies. At that concentration, there was no effect on macroinvertebrates in the channels. Reproduction of fish was not affected, but there was a small reduction in growth and increase in larval drift.

Crossland and Mitchell (1992) used artificial channels in Britain to test a mixture of chlorinated ethers from a petrochemical plant . Sublethal laboratory tests showed a single-species NOEC of 1.0 mg/L for growth of *Daphnia magna*. In the streams, the most sensitive effect was an NOEC of 0.44 mg/L for feeding of the crustacean *Gammarus pulex*. The difference of 2.4-fold should be considered reasonable agreement. Similarly in Oregon, five sublethal single-species laboratory tests of mortality, growth, and fecundity "adequately predicted the concentrations of diflubenzuron which affected ... stream communities" in artificial channels in a laboratory (Hansen and Garton, 1982). Effect was judged by biomass and diversity in the streams during five months of dosing. The insecticide reduced or extirpated many insects and

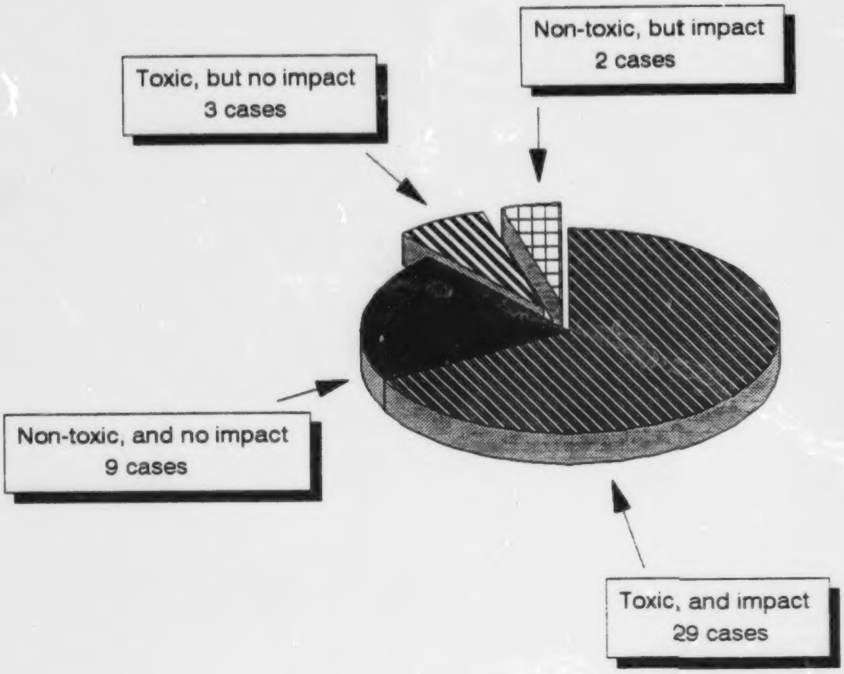

Figure 3 **Agreement of Predictions from Toxicity Tests on Effluents, with Observed Impact on Receiving-water Communities.** Data are for 43 river locations in North Carolina (data from Eagleson *et al.*, 1990). The two larger sections of the pie represent agreement, whereas the two smaller detached sections at the top represent disagreement of the observed and predicted status.

crustaceans at concentrations similar to effect-concentrations in the laboratory. It also impaired the decomposers and the sensitive benthic algae, both directly and indirectly by eliminating grazers so that only tolerant algae flourished.

Some validation has been attempted using mesocosms in ponds or enclosures in ponds. There was excellent agreement between laboratory and "field" when the pesticide atrazine was tested against eight species of algae. At concentrations of 100 to 155 μg/L, there was 50%inhibition of photosynthesis, respiration, and algal biomass in single-species laboratory tests, in microcosms, and in experimental ponds (Larsen *et al.*, 1986). Communities in enclosures in an Ohio pond were more sensitive to copper than the prediction from 7-d sublethal laboratory tests with *Ceriodaphnia* (Moore and Winner, 1989). Although the well-being of *Daphnia* and macroinvertebrates in enclosures was correctly predicted, there was a decline in populations of algae, rotifers, and some copepods. Tests of the insecticide Chlorpyrifos using *Daphnia pulex* gave conservative but reasonably accurate predictions of effects on populations of that daphnid. The EC50 for reproduction of *D. pulex* in conventional laboratory tests was 0.18 μg/L, while the EC50 for single-species populations in the laboratory was similar at 0.25 μg/L, and for populations held in enclosures within ditches used as mesocosms was 0.34 μg/L, within a factor of two of the conventional tests (van der Hoeven and Gerritsen, 1997). A multi-laboratory program compared sublethal tests in the laboratory with findings in artificial streams and pond enclosures, for copper, lindane, atrazine, and 3,4-dichloroaniline (EEC, 1992a). Sublethal thresholds in the laboratory were said to be protective of the outdoor communities for three of the toxicants, but the dichloroaniline caused effects at 1/10 of the lowest threshold in the laboratory.

See Section 6.3.2 for further discussion on mesocosms.

4.1.3 Literature Reviews

Some reviews of field validation have been published in recent years. Slooff *et al.* (1986) correlated single-species and multi-species tests, and the conclusions for sublethal effects are relevant here. They concluded that multi-species or "ecosystem testing does not lead to results that are dramatically different from those obtained with single-species tests". They calculated a predictive relationship for sublethal effects as follows.

$$\text{NOEC (ecosystems)} = 0.63 + 0.85 \log [\text{NOEC(single-species)}] \qquad (r = 0.85)$$

In a broad review of literature, Emans *et al.* (1991) found 17 comparisons that they considered reliable. A statistical analysis compared NOECs from multi-species studies with single-species tests on similar species. The relationship was linear and the authors concluded: "[T]here seems to be no reason to believe that organisms differ in sensitivity under field and laboratory conditions." They also concluded: "With reservations, due to this paucity of data, it is concluded that single-species toxicity data can be used to derive "safe" values for the aquatic ecosystem." [23]

Another review including validations described in Sections 4.1.1 and 4.1.2, reached similar conclusions. Among 63 individual comparisons by the originating authors, 53 predictions from sublethal tests on effluent agreed with effects found in communities, a rate of 84% agreement (Sprague, 1997). A more limited review of 13 of the same comparisons of field work with toxicity tests on surface water ("ambient tests") also found agreement of 84% (Parkhurst, 1996). The principles, approaches, and limitations of validation in aquatic environments is thoroughly reviewed in sequential chapters by Clements and Kiffney (1996) and Barbour *et al.* (1996).

4.1.4 Design of a Validation Study

A few publications provide guidance on designing validation studies in the field or in simulated communities such as mesocosms. Key items in the guidance may also be used to evaluate existing studies. Sanders (1985) gives excellent general advice on design of experiments for scientific validation of laboratory toxicity tests; Cairns (1986; 1988a) lists three desirable phases of such experiments. The OECD (1992b) presents good principles as recommendations from a workshop. A synopsis of advice is given in Sprague (1995), and the following key points are taken from that synopsis and from Sanders (1985).

Various degrees of scientific sophistication can be used for a field validation, but two important categories can be distinguished.

- In ***field evaluation*** the degree of correlation can be determined between the results of a single-species toxicity test in the laboratory and measurements of the quality of the receiving community. Most "field validations" in the literature fall into this category. The degree of correlation is assessed by a general and often subjective comparison.
- In ***scientific field validation*** there must be a higher degree of rigour in the design of the work, *i.e.*, a formal scientific experiment to determine whether the toxicity tests predict the effect on the community.

The scientific field validation uses the scientific method, with acceptability criteria set beforehand, so that the prediction of the toxicity test would either agree or disagree with the field results. The key steps in such a validation follow; some might be done in a different order, or overlap. If the formal scientific components were missing, the experiment would be considered a field evaluation rather than a scientific field validation. The process can be described as seven steps.

1. **Verify the toxicity test(s).** This simply means to confirm that the test operates as intended, seldom a problem with today's standard test procedures.

   One feature of utmost importance is that dilution water for the toxicity tests should be taken from "upstream" in the waterbody being studied. This usually requires some changes in the quality of culture water, to reduce the level of stress experienced by test organisms, because of the change in quality when they are introduced to control water from the site. For example, the pH and hardness of culture water could be adjusted to that of water from the site.

2. **Select the sites.** Sites must show promise of yielding legitimate scientific results. There must be a balanced community of aquatic organisms, which can be sampled with precision and relative ease, and there must be a comparable control area.

---

[23] The "safe" concentration was estimated by extrapolating downwards by methods described in Emans *et al.* (1991), from NOECs determined in single-species tests in the laboratory. That estimated "safe" concentration was compared with concentrations associated with effects or no effects in multi-species tests. The meaning of "safe" used herein is given in the Glossary.

The effluent should be relatively constant in nature, and it would be helpful if its toxic components were known and measurable by chemical tests. The discharge must cause an effect in the toxicity tests, and must also cause an observable effect on the receiving community. Ideally, there should be a gradation of effects in the waterbody, and a minimum of extraneous influences such as those from forestry or agriculture.

3. **Survey existing records** for useful background information on the site, particularly physical, chemical, and biological information.

4. **Initial screening and sampling program**. The main variables in the study should be assessed for feasibility. Exploratory sampling might be required on the magnitude and variation of wastewater toxicity, plume size in the receiving water, and composition of the biological community.

5. **Preliminary modelling/design/analysis.** This proceeds in concert with items 1 to 4.

6. **Design of detailed validation program**. Steps of the scientific method should be formally set up. (a) Formulate an hypothesis (e.g., "the community of organisms in the receiving-water will be deleteriously affected at concentrations of effluent greater than the IC25 in a sublethal toxicity test"). (b) Design the simplest possible experiment that will test the hypothesis. (c) Adopt the criteria that will be used to judge the results (e.g., what observation on the downstream aquatic community will be considered as a significant deleterious effect?). The statistical testing program must be designed ahead of time. The degree of precision must be decided, and hence the numbers of replicates. A sensitivity analysis, before the tests are done, might help to avoid indeterminate results. This technique is not commonly seen in pollution studies, and requires someone familiar with the process and capable of modelling.

7. **Carry out the program**. Samples of effluent or diffuse sources which reach the receiving water are collected. Tests are run, results analyzed, and the hypothesis is accepted or rejected. Results are reported.

4.2 Extrapolation of Toxicity Results

Key Guidance

- *Extrapolations are sometimes used to predict thresholds of sublethal effect from limited toxicity testing, perhaps from a few sublethal tests or perhaps from only a lethal test. Extrapolation methods range from simple to sophisticated, depending on data available and tolerance of uncertainty.*

- ***Application factors** are multipliers applied to LC50s to estimate presumed thresholds of sublethal effect in aquatic communities. Common values derived from experience are, for nonpersistent and noncumulative substances, 1/10 of the 96-h LC50 as a maximum at any time or place, and 1/20 or less as a 24-h average concentration after mixing. The corresponding values for cumulative toxicants are 1/20 and 1/100 of the LC50.*

- *The* **acute-to-chronic ratio (ACR)** *is, in essence, the inverse of an application factor. It is derived as the ratio of a measured acute LC50 to a measured sublethal threshold. Several tests may be combined for a more meaningful ratio. The ACR is commonly used today, and has the convenience of working with a ratio greater than unity. Once determined, ACRs might be used with discretion to predict to an unknown sublethal threshold in a different situation.*

- *If more complete toxicity data are available, extrapolation can be based on a function fitted mathematically to a distribution of species sensitivities. The "environmental concern level" (≈ NOEC) is the chemical concentration corresponding to a certain percentile (usually 95%) of the fitted distribution.*

- *Simple methods of predicting acceptable concentrations in the environment, based on a few data points, are usually satisfactory, and approximate the concentrations derived from more complex methods.*

Because of the reasonable correlation between measured toxicity and impairment of communities in surface waters (Section 4.1), toxicity tests are often used to predict no-effect concentrations in ecosystems. The extrapolations range from simplistic to sophisticated.

4.2.1 *Application Factors*

Application factors, as originally defined in the field of water pollution in the 1950s, were values less than unity, used to multiply an acute LC50 to predict "the concentration which will have no detrimental effect on aquatic life in the receiving water" (Henderson and Tarzwell, 1957). For example, an application factor of 1/10 or 0.1 was initially suggested as a multiplier of 48-h LC50s for the Ohio River Valley (ALAC/ORSANCO, 1955) [24].

When application factors were proposed in the 1950s, there were few laboratory estimates of sublethal thresholds for aquatic organisms, and experiments to determine thresholds were cumbersome and seldom attempted. There was somewhat more evidence from field studies, on concentrations of some chemicals and effluents that allowed unimpaired communities of organisms. The small amounts of information were used to estimate presumed reasonable application factors for a given species and a given wastewater. It was then assumed that the same factor (ratio) would be valid if the potency of the wastewater changed, or if there was a change in modifying conditions (e.g., binding and detoxifying agents). Quick determinations of LC50 could be made for the new effluent or modifying conditions, and multiplied by the application factor, to predict an environmental concentration without discernable sublethal effects. [25]

Application factors were intended to compensate for several variables that were unmeasured or difficult to ascertain: (1) the difference between the measured acute LC50 and a sublethal no-effect concentration in the receiving water; (2) greater sensitivity of other organisms in the receiving community; (3) undetected variation of the effluent of concern; and (4) deficiencies in the test method (Henderson, 1957; Henderson and Tarzwell, 1957).

At first, the numerical values assigned to application factors involved a fairly large dose of scientific intuition. Gradually, as results of using the factors were observed, and as sublethal laboratory experiments improved, the assigned values tended to become standardized (reviewed in Sprague, 1971). By 1964, the following values had become formalized in the USA for protection of aquatic organisms from toxicants whose true no-effect concentration was unknown (NTAC, 1968). The following application factors were used in the same way for many years in Ontario (OWRC, 1970; OMOE, 1984).

- Concentrations of materials that are nonpersistent ... or have noncumulative effects after mixing with the receiving waters should not exceed 1/10 of the 96-h LC50 at any time or place.
- The 24-h average of the concentration of these materials should not exceed 1/20 of the LC50 after mixing.
- For toxicants which are persistent or cumulative, the concentrations should not exceed 1/20 or 1/100 of the LC50 under the conditions previously described.

It was never recommended that fixed, "all-purpose" application factors be used indiscriminately, rather, they should be tempered with whatever information was available in a given situation. In the absence of specific information, however, application factors provided useful rules of thumb for their time, as scientists and technicians tackled problems of water pollution. They were intended primarily for appraising potential ecosystem damage from industrial effluents, or for estimating water quality objectives for new chemicals which had received little toxicological investigation (e.g., LC50 only).

---

[24] Neither an application factor nor an acute-to-chronic ratio were designed as a safety factor, and they should not be considered as such. Both are intended to be realistic representations of the ratio between the LC50 and a concentration that just fails to cause sublethal toxicity (i.e., ≈ the sublethal threshold, or ≈ the NOEC). A safety factor would "provide an extra margin of safety beyond the known or estimated sensitivities of aquatic organisms" (Stephan *et al.*, 1985). It was hoped that the application factor would predict a realistic threshold concentration for effects on an aquatic **community** (difficult to measure) from the lethal concentration for **one or a few species** (much easier and faster to measure).

[25] Application factors were applied to LC50s for other species, or other effluents, to meet practical needs in the absence of specific information. For example, if an application factor of 0.1 were found for an oil refinery effluent, it would be applied to other refinery effluents, and perhaps to other industries. A concrete example might be given. Say the LC50 of zinc for salmonids in soft water was 0.5 mg/L, and field and laboratory evidence indicated an NOEC of 0.05 mg/L. The application factor would be 0.05 ÷ 0.5 = 0.1 or 1/10. Now if the LC50 of salmonids in hard water were found to be 4 mg/L of zinc, then the sublethal threshold would be predicted as 4 × 0.1 = 0.4 mg/L.

Application factors continue to be seen in modern approaches, and in fact, the early values often seem reasonably in line with modern sublethal research. For example, Canadian guidelines give limited endorsement for use of application factors (CCME, 1991b) if there is not sufficient sublethal research to use the standard method of CCME (see Section 5.6). The CCME values are 0.05 for nonpersistent toxicants and 0.01 for persistent (and presumably bioaccumulative) ones, i.e., the traditional limits for 24-h averages listed previously. The CCME comments that these values have been accepted by some provinces (Ontario, Manitoba, Saskatchewan, British Columbia) and by the International Joint Commission (IJC). The CCME (1991b) specifies that factors must be selected from work that was designed for the purpose.

Based on the foregoing, it would seem that a factor could be judiciously transferred from work on one species or one toxicant to another, although that is not stated specifically by the CCME, nor are there specific directions on selecting the most appropriate factor among several. Modern values for application factors are also discussed in Sections 4.2.2, 4.2.3, and 4.2.4.

Some modern literature bends the meaning of application factor to serve particular situations that are entirely different from the traditional meaning outlined above. A reader encountering the term would be well advised to check the author's meaning.

4.2.2 Acute-to-chronic Ratios

These are essentially the same as application factors; their numerical value is simply the inverse of an application factor. Thus, an acute-to-chronic ratio (ACR) is the ratio between chemical concentrations exerting lethal versus sublethal toxic effects [26]. The term has gained favour in recent years because it uses a whole number to divide the LC50 and estimate the sublethal TOEC; most people find it easier to grasp the magnitude of a whole number than the decimal value less than unity used for application factors.

The term ACR is most frequently used when actual measured values are available for *both* the LC50 and the sublethal threshold, for a given species and toxicant. Following that, the calculated ratio could be used for the same predictive purposes as application factors, to predict TOEC under other (modifying) environmental conditions, for other similar industrial discharges, or with care, for other species.

As a broad generalization from experience with many substances and many toxicity tests, an acute-to-chronic ratio of 10 is considered to be a generally applicable, conservative estimate for use with industrial effluents and ordinary chemicals, when direct measurements are lacking (USEPA, 1991a; OECD, 1992b). This value is the same as that cited in Section 4.2.1 for maximum concentration of a toxic substance, but less restrictive than the value of 20 given for average concentration in the waterbody.

If results from several lethality tests were available for the same species, then the geometric mean should be used to characterize the findings [27]. Similarly, a geometric mean should characterize the results from several sublethal tests having the same endpoint (USEPA, 1991a; OECD, 1992b). If one sublethal effect (say the IC25 for reproduction) were lower than another (say the IC25 for growth), then the lower concentration would be used to calculate the ACR. If data for acute lethality were available for several similar species (e.g., different fishes) and sublethal data were available for only one of them, the acute-to-chronic ratio for that species could be applied to the acute data for the others (OECD, 1992b).

---

[26] This term started by using words that were reasonably accurate, but unfortunately became misnomers. Sublethal tests changed from life-cycle exposures that were *chronic* ("of long duration") to exposures of early life stages that were *acute* ("rapid"). ACR was intended to describe the ratio of lethal to sublethal endpoint (e.g., LC50 divided by IC25). Lethal tests are, indeed, almost always acute, but many modern sublethal tests, are also acute, e.g., the 7-d sublethal test with early life stages of the fathead minnow is clearly acute compared to the duration of life in that fish. A more extreme example would be the 20-min. exposure of sea urchin sperm to assess sublethal toxicity, compared to the life-span of years. Thoughtless misuse of "chronic" continues among aquatic toxicologists in North America. Because the term acute-to-chronic ratio (or acute-chronic ratio) is well established, it (and ACR) will be used here to some extent.

[27] The geometric mean is simply the antilog of the mean of the logarithms of the values (or the *n*th root of the product of *n* values). This useful mean will be similar in value to the arithmetic mean, for a compact and evenly distributed set of results. It is much less affected, however, by any extreme high value, which might be an erroneous outlier. In fact, the logarithmic value of endpoints such as LC50 should be regarded as the "true" representation, since endpoints are calculated as logarithms, from logarithms of the concentrations used in the test. The LC50 (or other endpoint) is customarily converted to an arithmetic value for convenience and comprehension by most people; it should be

4.2.3 Factors Based on Wider Data Bases

When evaluating a particular chemical or kind of effluent, a selection of relevant data on lethal and sublethal toxicity is often available. Attempts have been made to formulate some general factors that might be used with various sets of information, for predicting presumed thresholds of effect in the environment. Groups proposing these factors are cautious in stating what the predicted thresholds represent, usually designating them by names such as "potentially hazardous concentrations", *concern levels* (OECD, 1992b), or *protection levels* (Wagner and Lϕkke, 1991). The factors can be useful for flagging undesirable concentrations, but of course they would be subsidiary to any direct toxicological or environmental evidence.

A working group of toxicologists from several countries concluded that the following factors worked well for small sets of toxicological data (OECD, 1992b). The factors are intended to extrapolate to levels of "environmental concern".

Available Information	**Extrapolation Factor**
Lowest acute LC50 or EC50 within a data set of one or two species	1000
Lowest acute LC50, EC50, or IC50 in a set comprising at least one alga, crustacean, and fish	100
Lowest sublethal NOEC in a set comprising at least one alga, crustacean, and fish	10

These extrapolation factors decrease as the information base improves and the uncertainty of the toxicity estimate is reduced. Estimates based on chemical structure (QSARs) can be used in the hazard assessment, instead of experimentally determined endpoints (BKH Engineers, 1990; OECD, 1992b). These values are also given in Section 6 of a later guidance document (OECD, 1995a), along with a similar set of factors for extrapolating from toxicity tests done on fish-eating birds and mammals. The OECD documents (1992b; 1995a) use the wording "chronic NOEC" for the factor of 10, but clearly intended "sublethal" as discussed in Section 4.2.2, because the documents refer to short-term sublethal tests.

The rating based on extrapolation factors is very similar to one that is used by the USEPA Office of Pollution Prevention and Toxics *(sic.)*, which is responsible for preventing unreasonable risks to organisms due to the manufacture, import, processing, use, or disposal of industrial chemicals. That U.S. Office uses "assessment factors" for setting *concern levels* for new chemicals (Zeeman, 1995), as shown in the following table. (The words describing the categories has been slightly changed for clarity and accuracy.)

Available Data on Chemical or Analogue	**Assessment Factor**
Limited (e.g., only one acute LC50 estimated from structure-activity relationships)	1000
Base set of results for toxicity (e.g., acute LC50s for fish and daphnid, and algal IC50)	100
Thresholds of sublethal toxicity	10
Field testing	1

A similar set of factors had been adopted by the USEPA (1984b) for preliminary hazard assessments of chemicals. At that time, the EPA required five acute toxicity results, to use an extrapolation factor of 100, and recommended a factor of only three for wastewaters if the most sensitive sublethal data were used from three species (fish, invertebrate, and plant).

4.2.4 Statistical Extrapolations

If more complete toxicity data are available, extrapolation techniques can be based on distributions of sensitivity for

---

converted back to a logarithm before being used in mathematical procedures.

various species. These techniques involve fitting a function to the frequency distributions of the data. The environmental concern level is the chemical concentration corresponding to a certain percentile of the fitted distribution (usually 95%). The distribution-based methods assume that species in ecosystems have an approximately log-normal distribution of sensitivity toward any given toxicant, with a few very sensitive species, many moderately sensitive species, and a few very tolerant species. Reliable endpoints for at least eight species are required for the application of these methods (OECD, 1992b), and to avoid bias, they must represent different taxonomic and ecological groups. A workshop on extrapolation (OECD, 1992b) compared the available methods. Chapman *et al.* (1998) have provided a review of extrapolation methods with particular reference to ecological risk assessment.

The simplest extrapolation method was developed by the USEPA to derive water quality criteria to be used in regulation (Stephan *et al.*, 1985). The system was designed to protect 95% of exposed genera, and requires input of toxicity data from at least eight animal families, one plant, and in appropriate cases a bioconcentration factor and other data [28]. All the toxicity endpoints for a given species are used to calculate a geometric mean, and those in turn are used to calculate a geometric mean for each genus. Genus means are ranked, and the four lowest are used to calculate the chemical concentration closest to the fifth percentile of an assumed log-triangular distribution (a simple approximation of the log-normal distribution). Formulae for the calculation are given in Stephan *et al.* (1985) and OECD (1992b). The USEPA extrapolation method uses only the four lowest toxicity values, and so modest deviations from the upper part of the assumed distribution do not disturb the calculation. Endpoints characterized only as "greater than" some value (i.e., the highest concentration had no effect) can be included in the distribution. The concern level derived from the fifth percentile is intended to be protective of most species, most of the time. This formalizes an assumption that protection of 95% of the genera in an ecosystem will normally maintain ecosystem health. The EPA recommends, however, that stricter criteria should be developed when an ecologically, economically, or socially important species is at risk.

A mathematically more sophisticated extrapolation technique is based on a log-logistic distribution (Kooijman, 1987). It uses a series of chronic LC50s for a given toxicant to predict the median chronically lethal concentration to a most sensitive species. This assumes that the species are randomly distributed within the community for sensitivity. The estimated "concern level" has a specified probability that it will be lower than the chronic LC50 for the most sensitive species in the community. Kooijman's method estimates a chronic LC50, and so 50% of individuals in the most sensitive species might be expected to die, eventually, at the concern level, perhaps along with smaller numbers of less sensitive species.

Kooijman's concern level has confidence limits (usually 95%). They express the probability that the parameters of the model are estimated incorrectly and that the true probability of a more sensitive species occurring is different from that specified.

Kooijman's approach has been adapted to soil quality criteria (van Straalen and Denneman, 1991) and other variants of the method exist (Aldenberg and Slob, 1991; Wagner and Løkke, 1991). All these methods calculate the concern level as the geometric mean of LC50s or LOECs divided by an extrapolation factor derived from the estimated fifth percentile of the log-logistic or log-normal distributions. These methods parallel the approach of the USEPA, in adopting a goal of protecting 95% of the kinds of organisms in a community, not absolute protection of the most sensitive species. An easily followed description and integration of these methods is given in Section 5 of a guidance document (OECD, 1995a), although Canadian workers might do well to avoid some of the special terminology and acronyms coined there.

---

[28] The USEPA system is fairly complex but based on good postulates. The steps mentioned are followed for lethal ("acute") tests. The steps calculate a criterion for maximum 1-h exposure. If sublethal ("chronic") data were available for eight appropriate families of animals, they would be used in the same way to calculate a maximum 4-d average concentration. If sublethal data were limited, alternate procedures derive an acute-to-chronic ratio, to be used on the lethality criterion, to produce the 4-d criterion. If an estimated criterion was lower for plants, bioaccumulation, or "other", then that would take precedence over the criterion derived from eight families of animals. The choice of 1-h and 4-d time periods is supported in the document.

4.2.5 Comparison of Extrapolation Methods

Despite the differences in sophistication, the extrapolations described in Section 4.2.4 tend to produce similar results. Difference was within a factor of three, for six out of eight chemicals, when comparing a simple application factor of 100 on acute toxicity using three species (Table 3, column 1) with estimates from the USEPA method (column 4). Most estimates (51 of 56), for all the methods compared in Table 3, were within an order of magnitude of each other.

When NOECs were predicted by fixed extrapolation factors (columns 1 to 3 of Table 3), or by the method of Aldenberg and Slob (1991; column 5), most predicted NOECs were low by factors of 1 to 100, compared to actual observed effects on communities or species in microcosms (Okkerman *et al.*, 1990). Both methods, therefore, predicted greater toxicity than was actually found, so they apparently yield protective estimates for thresholds of sublethal effect.

In only three out of fifteen comparisons, the fixed extrapolation factors appreciably over- or underestimated observed thresholds. Fixed extrapolation could be used with limited data, but the distribution-based methods could not (Okkerman *et al.*, 1990).

The conclusion is that levels of environmental concern derived from only a few representative lethal or sublethal results are reasonable approximations of levels derived from more complete probability distributions (Table 3); they can be used with some confidence when data-sets are inadequate for the latter approaches

4.3 Toxic Units and Toxic Emissions

Key Guidance

- **Toxic Units (TU)** *express the degree of toxicity of an effluent, or the concentration of a substance, as a fraction of the measured endpoint of toxicity.*
- *Toxic Units are convenient because their magnitude parallels toxicity. This overcomes a conceptual difficulty with toxicity endpoints such as the LC50, which increase when toxicity decreases. Toxic Units are useful for calculations with effluent plumes.*
- *A* **Lethal Toxic Unit (LTU)** *is the actual or prevailing concentration of a particular toxic substance divided by the threshold LC50 of that substance (e.g., measured concentration in a river divided by LC50). For an effluent that is normally measured as percent concentration, the LTUs of the effluent are calculated as 100% divided by the LC50 of the effluent.*
- **Sublethal Toxic Units (SLTU)** *are calculated as the prevailing concentration divided by the IC25 or similar endpoint representing the sublethal threshold.*
- *Toxic units are dimensionless and can be added together, so they can serve as a base for predicting toxicity of mixtures of several chemicals or several effluents. However, LTU and SLTU cannot be mixed.*
- *The* ***amount*** *of toxicity discharged can be measured by the* **Toxicity Emission Rate (TER)**, *defined as TU multiplied by the volume of flow in* m^3/d. *TER governs impact in the receiving community.*
- *The* **Toxicity Emission Factor (TEF)** *is obtained by dividing the TER by the production rate of a factory [i.e., TEF = TU × (volume of flow in* m^3/t *of product)]. TEF is useful for equitable comparisons within and between industries, although it does not measure total environmental impact.*

As a chemical or effluent becomes more toxic, the LC50 or other endpoint becomes numerically smaller. If only lethal effects are considered then

$$\text{Toxicity} = \frac{1}{\text{LC50}}$$

This makes it practically and intuitively difficult to use test endpoints (LC50, IC25, etc.) in calculations such as dilution or the potential action of mixtures.

Table 3 Comparison of Extrapolations by Different Methods to Concern Levels for Eight Chemicals (OECD, 1992b). The first three columns represent customary extrapolation factors. The three species in Columns 1 and 2 were an alga, *Daphnia*, and a fish. Column 4 (Stephan *et al.*, 1985) is applied with a 50% confidence level so that the best estimate of the 95th percentile would have equal likelihood of over- and underestimation. Column 5 comes from Aldenberg and Slob (1991) applied at a 50% confidence level, Column 6 is the same but is applied at 95% confidence so that there is only a 5% likelihood of exceeding the true 95th percentile. Column 7 is the method from Wagner and Løkke (1991).

Chemical	1. Lowest LC50 or EC50, 3 species /100	2. Lowest NOEC, same 3 species /10	3. Lowest NOEC, 11 species /10	4. USEPA Log-triang. distrib. 95thprcntl. 50% conf.	5. Log-logistic distrib., at 50% confid.	6. Log-logistic distrib., at 95% confid.	7. Log-normal distrib., at 95% confid.
K_2CrO_7	0.01	0.01	0.01	0.085	0.71	0.010	0.013
NaBr	100	1	1.0	5.4	4.1	0.21	0.29
TPBS	0.16	0.1	0.032	0.27	0.27	0.042	0.052
2,4-DCA	0.013	0.0032	0.0032	0.019	0.064	0.0082	0.010
p-NT	0.075	0.1	0.032	0.19	0.45	0.10	0.12
DNOC	0.018	0.01	0.0032	0.019	0.031	0.0028	0.0037
Dimethoate	0.029	0.0032	0.0032	0.019	0.018	0.00036	0.00057
PCP	0.0047	0.0032	0.00032	0.0023	0.0043	0.00033	0.00045

4.3.1 The Toxic Units Concept

Extrapolations of toxicity from the laboratory to the environment can be simplified by calculating Toxic Units which are also more practical when handling toxicity data (Section 4.5). The USEPA (1991a), in its Technical Guidance Document, recommends using Toxic Units as an initial step in quantifying or modelling the toxicity of an effluent and specifying water quality criteria based on toxicity.

A Lethal Toxic Unit (LTU), as developed over some decades (reviewed in Sprague, 1970), is the prevailing (actual) concentration of a particular toxic substance, divided by the incipient lethal concentration for the substance in question. (An approximation such as the 96-h LC50 may be used for the denominator as an estimate of the incipient LC50, also called the threshold or time-independent LC50).

To characterize the toxic strength of an effluent, the "actual concentration" in the numerator would be 100%, because full-strength effluent is being dealt with, and

$$\text{Lethal Toxic Units (LTU)} = \frac{100\%}{\text{LC50 (in \%)}}$$

By that formula, if effluent A were "just lethal" (i.e., the LC50 was 100%), then LTU would equal 1.0 (the value is dimensionless)[29]. If effluent B had an LC50 of 5%, then LTU would be 100/5 = 20. It is easily recognized that effluent B is 20 times more toxic than effluent A.

When dealing with a chemical, other units of concentration would be used. If the concentration of a chemical were at 10 mg/L within the discharge plume, that would be the "actual" concentration used in the numerator to calculate Lethal Toxic Units. If the chemical had an LC50 of 2 mg/L for a particular species, that would be the denominator, and the toxic strength in that part of the plume would be 10/2 = 5 LTU of the chemical. Because Toxic Units were much greater than 1.0, strong lethal effects would be expected. The value of LTU also indicates the dilution necessary to reach a level that would be just toxic, in this case five-fold dilution. Values below 1.0 would predict non-lethality.

---

[29] For an effluent, it is not possible to calculate Lethal Toxic Units <1.0 because it is impossible to measure an LC50 >100%, i.e., greater than full-strength effluent. A chemical in the environment could have Toxic Units <1.0, because its LC50 can be measured by laboratory work, and its environmental concentration would usually be much less than that.

Sublethal toxic units can be similarly defined using an endpoint from sublethal toxicity tests as the denominator. Thus, for an effluent:

$$\text{Sublethal Toxic Units (SLTU)} = \frac{100\%}{\text{IC25 (in \%)}}$$

All calculations and uses of SLTU would parallel those outlined for LTU. For chemicals, the units might be mg/L in both numerator and denominator of the fraction, or alternatively they might both be μg/L. The numerator would be an actual concentration of interest, probably a measurement in surface waters. The denominator would be a laboratory determination of a threshold or near-threshold for sublethal effect, such as the IC25 or perhaps the TOEC. If surface water contained 10 μg/L of a given toxicant, and the IC25 for a sublethal test with fish was 2 μg/L, then SLTU would equal 10/2 = 5, and we would expect appreciable sublethal effects in the waterbody.

The LTU and SLTU are two different measuring units and cannot be mixed together in calculations. For modern concerns with pollution, Sublethal Toxic Units would be of more interest. In the USA, the two sets of units are usually called "acute" and "chronic" Toxic Units, terminology that is sometimes erroneous.

Certain other features should be mentioned. Any chemical units of concentration can be used, if they are the same in numerator and denominator for an individual substance. Once a value of TU is obtained, it can be compared with another obtained using different chemical units, because of the dimensionless character. The estimate of TU can be expected to change, however, with the test organism and endpoint used; therefore, organisms and endpoints should be standardized and reported within any given project.

4.3.2 Toxic Units and Effects of Mixtures

The toxicity of a mixture can be expressed as the sum of the TUs of each of the components, because TUs are dimensionless ratios. For example, a mixture containing five TU of one chemical and three TU of another contains eight TU in total. If the sum of the TU in the mixture equalled approximately 1.0 TU, then the mixture might be expected to produce the toxic effect (e.g., LC50) used as the base of the calculations. If the sum were <1, the effect might be expected to be correspondingly weaker, and if it were >1, a stronger effect might be expected.

Prediction of mixture toxicity by Toxic Units is predicated on the assumption that effects of the individual chemicals are additive. That assumption appears to be a conservative approximation of the truth, judging by an adequate body of sublethal studies in aquatic toxicity. Most chemicals with similar modes of action range from approximately additive to half-additive, i.e., sometimes contributing as little as half their potency to the joint effect. Those chemicals with divergent modes appear to contribute something like half of their individual toxicity (reviewed in Broderius, 1991; Christensen and Chen, 1991).

Examples can be given. Dutch workers tested mixtures of 5 to 25 organic substances having similar action, for effect on reproduction of *Daphnia* (Hermens *et al.*, 1984a). Mixtures were effective at 1.5 to 2 SLTU, that is, individual substances appeared to retain 1/2 to 2/3 of their potencies in the mixtures. Joint action in the mixture of 25 substances was strong (1.5 SLTU), even though each substance was present at only about 1/17 of its individual effective concentration. Recent studies continue to provide examples of near-additivity in mixtures of pesticides in laboratory tests (Bailey *et al.*, 1997), artificial streams (Carder and Hoagland, 1998), and in a river (Shigehisa and Shiraishi, 1998).

Comparable results were obtained for mixtures embracing several types of chemical structures and modes of action. A mixture of 14 diverse substances affected *Daphnia* reproduction at 1.9 SLTU (Hermens *et al.*, 1984b). Each of the dissimilar toxicants apparently contributed about half of its toxicity to the action of the mixture, when each was present at about 1/7 of its individual effective concentration.

Toxic units can be used to predict or evaluate the combined toxicity of effluents whose exact composition is unknown. Percent concentrations would be used in calculations, as previously described. One might predict the spatial extent and magnitude of expected biological effects in the receiving waterbody, based on the summed individual expected effects of concentrations in overlapping plumes. Several studies at lethal levels indicate near-additivity for effluents but unfortunately there is little documentation for effluents at sublethal levels. The sublethal pattern would be of much greater interest when considering protection of receiving communities.

There are some examples in aquatic toxicology of weak joint action, and even of independent action of chemicals in mixtures. More-than-additive effects are relatively uncommon, as are antagonistic effects (in which one toxicant would *reduce* the expected individual effect of another). There are several other models for evaluating toxicity of mixtures, mostly more complex or requiring specific knowledge of modes of action (reviewed in Christensen and Chen, 1991).

A parallel to the toxic units method exists in the generally accepted procedure for assessing toxicity of mixtures of **"chlorinated dioxins and furans"** (polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzofurans). There are 120 different isomers in this family of dioxins and furans. Toxicity of isomer mixtures is assessed by expressing the amount of each in terms of the amount of the most toxic isomer, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (2378-TCDD). That is, the amount (or concentration) of a given chemical is expressed in terms of its toxic concentration, with an additional factor to convert that to equivalents of 2378-TCDD. The calculation is basically the amount of a given isomer multiplied by the EC50 of 2378-TCDD, and divided by the EC50 of isomer. That is the same calculation as is used for toxic units with the added complication of using 2378-TCDD as the standard unit. In practice, a factor has been published for each isomer (NATO, 1988) to allow direct conversion of a given amount of isomer to the base unit. It is standard practice in toxicology to assume additivity of the 120 isomers. The isomers are similar in structure and have similar toxic actions, but the parallel is nevertheless encouraging for use of the toxic units concept in environmental toxicology.

4.3.3 Mass Loading of Toxicity

The importance of an effluent to the receiving environment depends not only on its toxic concentrations, but on the total **amount** of toxic substance, taking into account the volume of the discharge. Effluent streams of different volumes and toxicities can be rationally compared in terms of *Toxicity Emission Rate* (TER), defined (California, 1972) as:

Toxic Emission Rate = TU × (volume of flow in m^3/d)

Expressing the effluent flow as the TER essentially normalizes the effluent volume for the toxicity it contains. The tool is useful for evaluating relative amounts of toxicity contributed by various industries on one waterbody, or by various component streams within one industrial operation. The TER also provides a useful second way (in addition to conventional toxicity tests) of assessing changes in effluents such as reduced discharge of process water. The effluent might appear to worsen according to toxicity tests, but the TER might show reduced toxic loading [30].

A further refinement of the toxicity emission concept allows comparison between the processes at different industrial sites by expressing the amount of toxicity in terms of production, e.g., per tonne of industrial goods produced.

Toxic Emission Factor = TU × (volume of flow in m^3/t of pulp)

The concept of TEF is not unusual; current regulations use a "technology-based approach" to set physicochemical limits for some traditional pollutants. Canadian pulp and paper regulations, for example, restrict BOD discharge to 7.5 kg per tonne of product. The TEF follows a parallel approach by incorporating units of production [31]. The TER deals with the amount of effluent per day, and

---

[30] Suppose a mining operation considered technology to reduce its water use by a factor of four, desirable for water conservation. Suppose also, that the new process gave a two-fold reduction in discharge of toxic substances. Despite the improvements, the concentration-based toxicity of the effluent would double; it might change from non-lethal to lethal. Environmentally, it would be reasonable to focus on the reduction in loading, rather than the increased *concentration* of toxicants in the effluent.

[31] If a factory had a TER of 100, the units of TER would customarily be expressed as "100 m^3/d", although a more proper designation might be "100 m^3/d of LTUs" (or SLTUs). An easy way to think of this would be that the factory would release the equivalent of 100 m^3 of "just-barely-lethal" effluent every day. The actual discharge might be 10 m^3/d of effluent with LTU = 10, or 50 m^3/d of effluent with LTU = 2. Similarly, TEF is customarily expressed as "m^3/t", although it might be more properly signified as "m^3 of LTUs per tonne of product".

TEF is a fair way of comparing industrial plants that use different processes, because it expresses quantity of toxicity, corrected for factors of plant size and water use efficiency. The TEF can also compare discharges at one operation over the years, despite changes in production or processes.

Calculations of TER and TEF have been used within industrial plants, especially pulp mills, for discovering which processes contributed the most toxicity. Scroggins (1986) was able to rate the important sources within a bleached kraft mill in Ontario (water from washing and screening, blow heat accumulator, and bleach plant). Federal government biologists proposed TEF for use at pulp mills (Wilson *et al.*, 1975), and industries have used TER and TEF for some time in government or internal reports (e.g., Naish, 1975; Beak, 1979; Fahmy and Schiefer, 1980).

A final example follows on the usefulness of considering *both* toxic concentration and toxic loading at industrial operations. Data on LC50, LTU, TER, and TEF, are for 1987 discharges from two Ontario paper mills of similar size and similar industrial operations (Sprague *et al.*, 1991; Appendix A.)

	Mill A	Mill B
LC50, % effluent	29	10
Tonnes of product/day	284	269
Effluent flow, m^3/d	3 359	11 928
Effluent flow, m^3/tonne of product	12	44
Lethal Toxic Units	3.4	10
Toxicity Emission Rate, m^3/d	11 420	119 280
Toxicity Emission Factor, m^3/t	41	440

From the LC50s and TU, the effluent of Mill B is about three times as toxic as that of Mill A. However, calculations of TER and TEF show that Mill B is discharging about ten times as much toxicity to the receiving water, and about ten times as much per tonne of paper produced. In this case, TER and TEF show a similar ratio between the two mills because the mills are of similar size; the two measurements need not be similar, and they might sometimes convey two distinct messages.

4.4 The "Weight-of-evidence" Approach

Key Guidance

- *Use of three basic investigative techniques (chemical measurements, toxicological tests, and biological surveys) allows a "weight-of-evidence" interpretation of environmental impact. The case for detrimental effects on a given ecosystem is built on collaborative evidence from each of the three techniques.*

- *This does not mean that all possible tests should be applied in every project; each has a specific purpose and role. Chemical measurements and toxicity tests can be used to predict effects in the environment, while biological surveys supply the authoritative audit. Accordingly, results from the three techniques cannot be expected to be equally revealing.*

- *The Sediment Quality Triad formalizes the use and interpretation of this trio of techniques. The pattern of effect can indicate the nature of environmental impact (Table 4).*

In current technical writing on environmental pollution, the "weight-of-evidence" approach simply means use of information from all sources, but particularly from three techniques, a "triad" of (a) toxicity tests, (b) chemical measurements, and (c) biological surveys in the field. These three approaches are compatible, and together they reveal more about the presence, strength, nature, and effects of toxicants than could be discovered from any single approach (Section 1.2). Each technique serves to both confirm and extend the understanding of environmental contamination in surface waters, sediments, or soil.

At the same time, results from the three techniques must not be blindly rolled together in the hope that a clear picture will somehow emerge. Munkittrick and McCarty (1995) caution: "The weight of evidence approach often attributes equal strength to a wide variety of responses, without a close examination of their scientific applicability or validity." Each approach differs in function and should be used when there is a specific purpose for which it is

---

ignores the amount of goods or product which the industry manufactures. Conversely, TEF is based on some unit of goods production, but ignores the time unit, such as "per day". However, if the amount of goods production per day is known, one can move from TER to TEF and vice versa. If a paper mill had TER = 200 000 m^3/d, and produced 1000 tonnes of paper per day, then TEF would be 200 000/1000 = 200 m^3/t.

suitable. The roles of each technique are outlined in Section 1.2, but are repeated briefly here.

- *Toxicity tests* tell **whether** a toxic effect is caused by a sample of chemical, waste, surface water, sediment, or soil. They cannot by themselves determine the identity of the toxicant.
- *Chemical tests* can identify **what** is the likely cause of toxicity, and its concentration. They can predict whether a toxic effect is likely, but cannot give proof of toxicity.
- *Biological surveys* of the receiving environment provide the **confirmatory audit** of whether or not there has been an effect on the living community. (Excellent introductions on methods for monitoring benthic invertebrates are provided by Hellawell (1978), Merritt and Cummins (1984), or Rosenberg and Resh (1993). A technical guidance document (EC, 1997e) outlines methods and required samples for Environment Canada's monitoring program for the pulp and paper industry).

Another triad of tactics is given in the following categorization and helps demonstrate the purposes of the available techniques (Section 1.2 and Sprague, 1991).

- (a) *Tests of effluent quality.* This tactic compares test results with any goals that have been adopted for good industrial practice, e.g., no lethality for fish, or pH control within the range x.x–y.y. The findings do not necessarily have a direct relationship to quality of the receiving environment, but identify discharges that conform to good industrial practice.
- (b) *Tests of quality in the receiving water, soil, or sediment.* This second tactic uses direct chemical and/or toxicity measurements in the environment to determine whether guidelines of environmental quality are met. (In practice, measurements might be made on an effluent, followed by calculations to estimate the levels in the environment after dilution.) This is a predictive approach because the guidelines used for comparison represent only the best judgement of scientists, and single-species toxicity might differ from effects in a community. Nevertheless, direct testing of the receiving media is a powerful tactic. It has been relatively neglected in Canada, but chemical and toxicity tests of sediment are now coming into much greater use.
- (c) *Biological surveys in the receiving community.* This third tactic provides the over-riding "proof of the pudding" on whether predictions about effects or lack of effects are confirmed in the real community of organisms.

Any one of these approaches might be used first, depending on the local situation. In some instances, a field survey of the biota might be a logical way to start a program—if it showed no effect of a discharge, there might be no point in chemical or toxicological measurements. More likely, an economical way to start any investigation would be to use chemical or toxicological tests.

A particularly good example of combining chemical and biological techniques of toxicological investigation is in toxicity identification procedures (TIE, Section 4.7).

4.4.1 Sediment Quality Triad

The complementary use of toxicity, biological monitoring, and environmental chemistry applies equally well to water, soil, and sediment, but for sediment it has been formalized as the *Sediment Quality Triad* (Chapman, 1986; 1990; 1996).

In the sediment triad approach, sites are studied by (a) monitoring benthic invertebrates, (b) environmental chemistry, and (c) a battery of toxicity tests on whole sediment, porewater, elutriates, or extracts. Results are compared with parallel measurements at nearby reference sites. Results for a given polluted station can also be summarized as *ratio-to-reference values*, the ratio of a measurement at the test site to that at the reference site. The ratios can be presented on a tri-axial graph, with confidence limits (Alden, 1992; Canfield *et al.*, 1994). As an alternative, data can be scaled from zero (reference site) to 100 (worst site) and presented as triangles on axes with those scales (Figure 4). Other examples are presented in Chapman (1986) and Chapman *et al.* (1987; 1991).

Methods of interpreting and applying the Sediment Quality Triad have been relatively simplistic and to some extent subjective. There has been criticism of the lack of power and understanding from separate calculations of effect for each of the three categories, and from expressing the effects as indices. Green *et al.* (1993) make those comments and show various techniques of multivariate statistics, to allow for interrelationships and to better *understand* what effects are taking place. They point out that without multivariate analyses, there is considerable

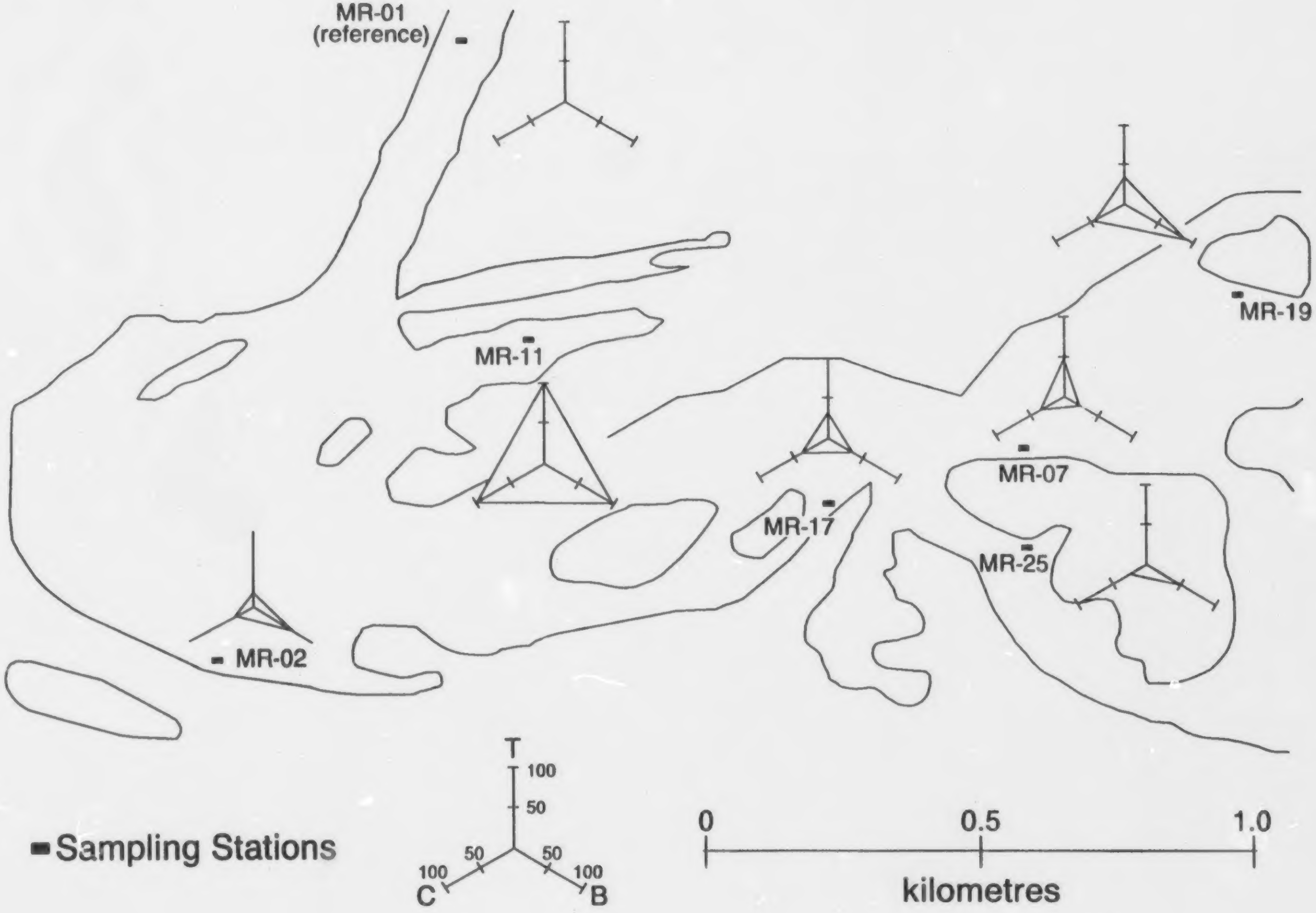
MR-01
(reference)
MR-11
MR-19
MR-07
MR-17
MR-25
MR-02
Sampling Stations
T
100
50
50
50
100
100
C
B
0
0.5
1.0
kilometres

Figure 4 Use of Tri-axial Diagrams to Illustrate Results from the Sediment Quality Triad Applied to Metal-contaminated Sediments (Canfield *et al.*, 1994). T, C, and B represent toxicity, chemistry, and biological survey. Axes are scaled from zero (reference value) to 100 (worst value measured).

concern of confounding natural environment gradients with pollutional effects, "and false conclusions could be reached". Methods of interpreting data from a triad have started to improve, and Chapman (1996) provides examples of the advantages of multivariate analysis and presentation, compared to the relatively simplistic triangular diagrams.

The three facets of the sediment quality triad might not always be equally revealing, indicating a complex situation that might deserve further investigation. The different axes of the graph might be unequal, because they measure very different things and have different purposes and sensitivities.

Table 4 should assist in interpreting different combinations of responses for the sediment quality triad. The same interpretations should be useful in assessing any medium (water, sediment, or soil).

An assessment of contaminated sediments near three Canadian mines demonstrated the strength of the interpretive approach based on the sediment quality triad. The correlations were quantified, between chemical characteristics and toxicity of the sediment, between chemical characteristics and the benthic community, and between toxicity of sediment and the benthic community. The assessment used principal components analysis and multiple correlation coefficients using the principal-component variables. At two of the mines, correlations were significant for all three of the comparisons. Sediment toxicity appeared to be closely related to availability of metals, while the benthic community seemed partly responsive to changes in physical habitat. At the third mine site there were only weak correlations among the triad components. The analysis suggested that the elevated metals in the sediments were not easily available to the benthic invertebrates, nor to the benthic invertebrates (AETE, 1998a).

4.5 Toxicity Indices

Key Guidance

- *Toxicity indices express the results from several different toxicity tests as a single number which rates the toxicity of the sample.*
- *No fixed rules have been established for designing toxicity indices, but design decisions will emphasize one or more aspects of the testing.*
- *The PEEP index is currently used for the St. Lawrence River Action Plan, to rank industrial effluents according to the input from four toxicity tests.*
- *Any kind of toxicity index should be used circumspectly because of the loss of information.*

Water quality indices based solely on *chemical and physical* measurements have been available for several decades, e.g., a design suggested for Canada by Inhaber (1975). Usually, indices provide a score for each variable (e.g., pH, or metal concentration) by a comparison with recognized water quality guidelines. The individual scores are then combined into a total or average index that is used as a convenient measurement of overall water quality.

Chemical indices continue to have great appeal because their single value is easily grasped. However they could sometimes be deceptive. An index could encompass a dozen or so measurements that provided a reasonable assessment of common kinds of pollution (Prati *et al.*, 1971), but might fail in a given situation because it did not include an important toxicant[32].

There would be similar appeal for use of a toxicity index that integrated diverse biological and physicochemical measurements, but such indices are not common. In Canada, the PEEP index incorporates toxicity (see following text, and Appendix B). Another Canadian index combines results from toxicity, biochemical assays, and microbial enumerations from both water and sediments; it was used along the St. John River (Dutka *et al.*, 1988). A proposed screening method attempts to calculate overall

---

[32] The index of Prati *et al.* (1971) includes pH, oxygen, BOD, COD, permanganate, suspended solids, ammonia, nitrate, chloride, iron, manganese, ABS surfactant, and CCE. That would fail to assess a spectrum of subtle and dangerous toxicants such as PAHs, PCBs, and chlorinated dioxins. It would even fail to assess pollution by copper and zinc, often found in Canada without accompanying major changes in those variables included in the index.

Table 4 **Interpretation of Effects in the Sediment Quality Triad (from Chapman *et al.*, 1992).** An effect is shown as positive if it is significantly different from the condition at the reference site. No difference is indicated as a negative.

Contamination	Toxicity	Community Alteration	Possible Conclusions
+	+	+	Strong evidence for pollution-induced degradation
-	-	-	Strong evidence, there is no pollution-induced degradation
+	-	-	Contaminants are not bioavailable
-	+	-	Unmeasured chemicals or conditions exist with potential to cause degradation
-	-	+	Alteration is not due to toxic chemicals
+	+	-	Toxic chemicals are stressing the system
-	+	+	Unmeasured toxic chemicals are causing degradation
+	-	+	Chemicals are not bioavailable or alteration is not due to toxic chemicals

environmental impact by incorporating into the index, the amount released into the environment, persistence, and bioaccumulation, as well as known toxicity to organisms including humans (Swanson *et al.*, 1997).

A precursor to a toxicity index is the common use of tests to rank environmental samples or hazardous sites according to the severity of toxicity. Ranking often determines priority for action on the most toxic effluents or contaminated sites. Ranking the samples can become complicated if toxicity has been measured with several tests that produced variations in rank. Collapsing the responses into a *toxicity index* would express hazard in a single number.

No fixed conventions have yet appeared for design of a toxicity index, but some general considerations are indicated herein.

An index based on the sum or average of toxicity endpoints is the simplest to devise. It might be desired to combine tests of acute lethality with sublethal tests in order to include a spectrum of organisms. A deliberate bias toward either of those categories could be built in by the kinds and numbers of tests selected and included, or by multiplying results of one category by a weighting factor.

Indices are easier to construct if toxicity endpoints are first translated into Toxic Units (TU). The numerical values can then be summed like chemical properties of a sample (Section 4.3.2). An alternative would be to classify results on an ordinal scale (e.g., 0 to 10) based on the observed severity of effect. The approach is more subjective, but at least it incorporates expert judgement that should enter the assessment of data at some point. A ranking scale also allows any kind of environmental measurement to be included in the index which could be detrimental as well as beneficial.

In the design of an index, two decisions must be made about placing weights on components. The first is whether any endpoint (e.g., a test with a particular species) is to be considered more important and thus weighted more heavily. The second decision is whether the extent of toxic response (e.g., number of tests showing toxicity) is more important than the strength of effect (potency to one or more species), or vice versa.

The PEEP index was developed and is currently used under the St. Lawrence River Action Plan for ranking industrial effluents (Costan *et al.*, 1993). Four toxicity tests are used: algal growth, SOS-Chromotest for genotoxicity, bacterial luminescence, and the sublethal *Ceriodaphnia* test. Some of the tests are used before and after aeration of the sample, and some produce two endpoints, so that 10 endpoints are finally combined. The toxicity assessment is combined with a measure of the volume of discharge, to express as a single number, the potential for causing biological effects. The index normally falls between zero and 10, and the complete procedure is described in Appendix B. The PEEP index purposely weights in favour of sublethal toxicity, and considers toxicity to several species to be more serious than strong toxicity to one species (Blaise *et al.*, 1985). The index was tried as part of the Fraser River Action Plan in British Columbia, resulting in a suggestion that a simplified version might be useful for the Fraser River.

Chimiotox is another rating system developed for the St. Lawrence River Action Plan as a companion to the PEEP index. It is not a toxicity index in the sense used here because toxicity is not actually measured. Chemical profiles of municipal and industrial effluents and known toxicity information are used to rank the effluents for potential environmental damage. The daily load of each chemical component is multiplied by a toxicity factor, and the sum of toxic loads produces the rating (see Appendix B).

Toxicity indices should be used cautiously because there is a loss of information in the calculation of one summary value. Results of individual tests should be examined for questionable results and for exceptional endpoints that might skew the index.

4.6 Structure-activity Relationships

Key Guidance

- *Structure-activity Relationships (SARs) predict the toxicity of organic compounds based on physical and chemical properties.*
- *A qualitative SAR predicts approximate toxicity by comparing compounds with other compounds of similar structure.*
- *Quantitative structure-activity relationships (QSARs) are regression equations relating toxicological endpoints to physicochemical properties within a class of compounds. Many QSARs are based on the octanol–water partition coefficient (K_{OW}).*
- *Different QSARs are required for different classes of chemicals.*
- *SARs are used as one means of evaluating new chemicals for use.*

Most toxic chemicals must move from the environment into the organism, to exert their detrimental action within the cells. Both processes depend on the elemental composition and structure of the chemical. Accordingly, the toxicity of organic chemicals can be predicted from their physicochemical properties by *Structure-activity Relationships* (SARs). There are diverse uses of SAR.

- **Predicting the effect of new commercial chemicals**. Production, importation, and use of new chemicals are subject to regulation in Canada, USA, and Europe. SARs play a role in estimating the degree of toxicity during a review period.
- **Setting priorities**. Of the tens of thousands of organic chemicals in production, most are probably not highly toxic, but also, most have not received adequate assessment of toxicity. SARs can be used to rank chemicals for priority in testing.
- **Developing preliminary water quality guidelines**. SARs can be used to estimate likely thresholds of toxic effect in surface waters, if toxicity information is lacking.
- **Rapid estimates of missing toxicity data**. SARs can quickly estimate toxicity when there is no time to perform a test (e.g., a chemical spill).
- **Guiding tests**. Predictions of toxicity by SARs could improve efficiency of testing or even reduce the number of tests required. Optimum ranges of test concentration could be chosen. If several chemicals were candidates for a particular commercial purpose, SARs could indicate the less toxic ones, reducing the need for costly screening. A SAR might suggest that the LC50 of a chemical would be in excess of its solubility, and consequently, effort might be diverted to sublethal tests. If measured toxicity were much higher than predicted, the test should be checked for anomalous conditions, and/or the SAR should be examined for reasonableness.

4.6.1 Types of Structure-activity Relationships. *Qualitative SARs* represent the simplest approach. The toxicity of an untested compound can be predicted by comparison with other similar compounds. For example, if all the tested alcohols of a particular group were toxic to fish at concentrations above 100 mg/L, a structurally similar new alcohol would predictably be toxic in the same range. The nearest analogue to an untested chemical, or the two analogues having characteristics that bracket it, can be used to predict its toxicity, if the usual physicochemical properties do not correlate well with toxicity (Nabholz *et al.*, 1993).

Qualitative SARs (sometimes called "read-across" estimates) can be useful to screen out compounds with a probable toxicity so low that toxicity testing would be superfluous. The reverse application is to identify compounds that are probably toxic. Qualitative SARs are not, however, reliable for predicting potency among toxic compounds, because relatively minor structural changes can have large influences on toxicity. Predictions of qualitative SARs are best restricted to chemicals that differ in only minor ways from other compounds, for example, addition of one monomeric unit (such as a methyl group) to a long-chain polymer, or removal of one.

Quantitative structure-activity relationships (QSARs) are regression equations relating toxicological endpoints to physicochemical properties of a given class of organic chemicals. Endpoints are acute LC50s or sublethal IC50s, EC50s, or NOECs. Physicochemical properties are diverse and include hydrophobic quality, electronic descriptors such as dielectric constant and ionization potential, and steric qualities such as molar volume and total surface area (OECD, 1992a).

QSARs have been extensively developed for aquatic systems because of the large database and the relative simplicity of quantifying bioavailability. QSARS have also been developed for terrestrial wildlife, based on classical oral or inhaled dosages. QSARs for sediments and soil are being developed (Van Gestel and Ma, 1993) sometimes following the principles used for aquatic QSARs.

In the *regression approach* to QSARs, a simple or multiple regression equation describes the LC50 or other endpoint as a linear function of a physicochemical property or properties of chemicals in a given class (e.g., phenols). Ideally, toxicity would be measured with one species and one endpoint under identical conditions, although this is not always achievable. Toxicity is expressed in mol/L to eliminate molecular mass, and endpoints are log-transformed to achieve linearity. Ordinary statistical techniques such as least-squares regressions are used.

A second approach to developing QSARs uses a basic toxic contribution from the simplest parent compound, then adds contributions from each substituent group (NH_2, CH_3, OH, etc.), as determined by multiple regressions on a large number of compounds. This has been done for 105 substituted benzenes (Hall *et al.*, 1989).

By far the most common chemical descriptor in QSARs is valid for some organic chemicals, but not inorganic ones. This is the *octanol–water partition coefficient* (K_{ow} or P_{oct}), useful because it is a strong predictor of the tendency of a compound to be accumulated by organisms (Veith *et al.*, 1979). The K_{ow} describes the equilibrium distribution of a chemical in octanol (a nonpolar solvent) compared with water (a polar solvent). The K_{ow} therefore describes the chemical's tendency to partition into the lipids of cell membranes instead of remaining in the water. Hydrophobic compounds, with a high K_{ow}, will tend to migrate into the fats and lipids of tissues and accumulate there, causing toxicity (Bertelsen *et al.*, 1998). Partition coefficients are generally expressed on a logarithmic scale, and the range for most biologically active organic compounds is from -5 to +6, although values greater than +8 are possible. Further discussion is given in Section 6.4.

Different QSARs are required for different classes of chemicals according to their physicochemical attributes, or to their mode of action. Mode of action has a stronger basis in toxicological theory, but often it is not fully understood (OECD, 1992a). Some modes currently recognized include narcosis, interference with energy metabolism (respiratory uncoupling), interference with the nervous system (acetylcholinesterase inhibition), and respiratory irritation (McKim *et al.*, 1987).

4.6.2 Narcotic and Non-narcotic Models

Narcotics are the largest class of toxic organic chemicals, and have the most extensively developed QSARs. Narcosis is defined as a generalized physiological decline caused by an organic chemical. It is reversible, and is considered a baseline effect that is not caused by specific elements (e.g., chlorine) or chemical structures (e.g., benzene rings) (OECD, 1992a). This non-specificity explains why K_{ow} alone can be an adequate predictor of toxicity caused by external concentrations [33]. In principle, any organic compound can act as a narcotic, but the category includes aliphatic and aromatic hydrocarbons, chlorinated hydrocarbons, alcohols, ethers, ketones, weak acids and bases, and some aliphatic nitro compounds (Veith *et al.*, 1983). The OECD (1995a) provides descriptions and lists of five classes of chemical according to their action, starting with chemicals that act by narcosis, and also gives numerical tabulations of QSARs for such chemicals.

---

[33] The relationship also predicts that molar concentrations of narcotics within intoxicated organisms should be similar for all narcotics. This is a very useful feature for predicting toxicity. It also helps to explain the near-additive joint action of narcotics (McCarty *et al.*, 1992; McCarty and Mackay, 1993; see Section 6.4).

An example of a QSAR for chemicals that act by narcosis is:

$$\log 1/C = A \log K_{ow} + B$$

where: C is the LC50 or other endpoint, and
A and B are constants of the linear relationship.

A clear-cut and simple example of an excellent QSAR is seen in Figure 5, for a series of alcohols.

Other examples for narcotic effects on fish are shown in Table 5, where inspection of the equations will show both similarities and differences. Values of A are normally around 1; values for B are about 1.0–1.5 for lethality and 2.0 for sublethal endpoints (OECD, 1992a; Bol *et al.*, 1993a; 1993b).

Many other listings of QSARs are available, most notably the ECOSAR program with a computer disk containing 100 SARs for 40 chemical classes (e.g., four classes are for categories of surfactants) (USEPA, 1994c).

For soil organisms, QSARs are also being developed from the lipid–water partition coefficient (Van Gestel and Ma, 1993). Once again, this is a feature that is useful for organic chemicals, including some metal-organic complexes such as methylmercury. The assumption is that toxic chemicals move into organisms from solution (in the soil porewater), as in aquatic systems. The tendency to bind with soil particles or move into soil solution is also a function of K_{ow} (see Section 4.11.3). The same principle applies to chemicals in sediments.

Chemicals that act through a specific toxic mechanism are decidedly more toxic than narcotics. They can be recognized by the deviation from QSARs developed for narcotics (Ikemoto *et al.*, 1992), a deviation called "excess toxicity" (Nabholz *et al.*, 1993) that indicates the need for a more specific QSAR.

QSARs for non-narcotic chemicals are less developed, but the existing relationships are generally based on homologous groups of chemicals, rather than on common modes of action. The Netherlands divides non-narcotics into four classes for hazard assessments: phenolic compounds; reactive chemicals; specifically acting chemicals; and unclassified "others". For the first two categories, thresholds are predicted with extrapolation factors. Predictions for specifically acting and unclassified chemicals have not been made (Verhaar *et al.*, 1992).

4.6.3 Use of Structure-activity Relationships

In Canada, new chemicals are controlled under the Canadian Environmental Protection Act (CEPA). The New Substances Notification Regulations promulgated under CEPA require producers of new chemicals to provide toxicological data. SARs might be used, when appropriate, as a substitute for experimental data (see Section 5.7).

In Europe, SARs are coming into use for hazard assessment of new chemicals under the European Economic Community (EEC) (Bol *et al.*, 1993a). The Netherlands is taking the lead on behalf of the EEC, and uses SARs as an aid in setting environmental quality criteria (Verhaar *et al.*, 1992; Bol *et al.*, 1993a).

In the United States, the USEPA is limited to a period of 90 days for review of new chemicals before they come onto the market, and has reviewed some 26 000 chemicals since 1979. QSARs are routinely used in predicting aquatic toxicity quickly (program reviewed by Zeeman, 1995). The USEPA has developed QSARs for all new commercial chemicals evaluated since 1981, and has about 300 QSARs for some 100 classes of compounds, most of which are based on K_{ow} and molecular weight. Many QSARs are available on computer disk (USEPA, 1994c).

A common goal for accuracy is that predictions from QSARs should match laboratory results within an order of magnitude. Nabholz *et al.* (1993) found that predictions met that goal for 85% of 920 measured endpoints for 462 chemicals; toxicity was overestimated for 9% of the predictions and underestimated for 6%. The EEC tested USEPA QSARs for 130 organic compounds; 77% of the predictions for fish agreed with test results within a factor of five, and 67% of predictions for *Daphnia* agreed similarly (USEPA and EEC, 1993).

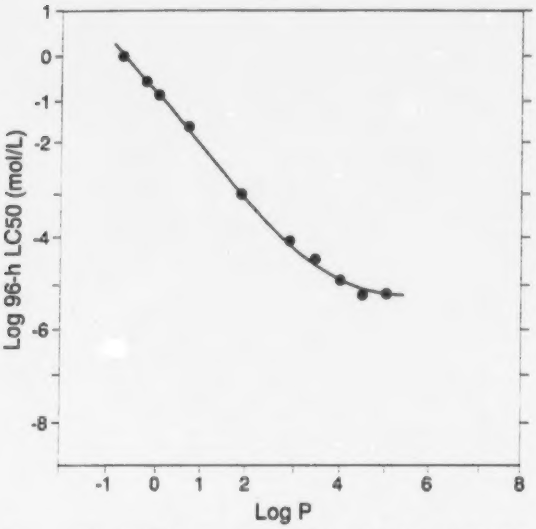

Figure 5 A Quantitative Structure-activity Relationship (QSAR) for Aliphatic Alcohols. The 96-h LC50s for fathead minnows are related to the octanol–water partition coefficient (P). Modified from Veith *et al.* (1983).

Deciding on the best QSAR for a particular application requires expert judgement based on knowledge of organic chemistry. The assignment of chemicals to the appropriate class is one of the largest sources of uncertainty in the use of QSARs (OECD, 1992a; USEPA and EEC, 1993). Even the menu-driven computer program of ECOSAR requires some expert knowledge of toxicology and organic chemistry (USEPA, 1994c). There is a trend developing toward establishing one validated, broadly based QSAR per endpoint within each chemical class, which would then be used for all toxicity evaluations.

QSAR is a very active field of research. An excellent and learned review has been provided recently by Lipnick (1995) and somewhat earlier reviews are in Tosato *et al.* (1990) and in a series of papers from a symposium, edited by Elzerman (1992).

4.7 Toxicity Identification and Reduction

Key Guidance

- *Toxicity Identification Evaluation (TIE) identifies the toxic component(s) of a mixture. In Phase I, a sample is manipulated physically and chemically to isolate chemical fractions. Toxicity tests before and after each manipulation identify the toxic fractions and guides the chemical analysis of Phase II. In Phase III, results of the preceding two phases are checked by various techniques.*
- *Some possible manipulations are aeration, filtration, chelation of metals, oxidant reduction, pH adjustment and solid-phase extraction. A fundamental tactic is to remove the dominant toxicant, then search for others. The general procedures should be adapted to the particular sample.*
- *TIE can be applied to any medium, but methods for soils and sediments are in developmental stages.*
- *Toxicity Reduction Evaluation (TRE) is similar, but focuses on finding suitable methods for waste treatment.*

A formal set of methods was developed in the USA, to systematically isolate and identify the toxic component(s) in a mixture, using both chemistry and toxicity tests (USEPA, 1989a; 1989b; 1991c; d; 1996). This set of methods is now used widely, and the name *Toxicity Identification Evaluation* (TIE) is usually adopted despite its redundancy and uncertain grammatical construction. The methods are useful for any complex mixture, be it an industrial effluent, surface water, sediment porewater, or hazardous waste leachate (Norberg-King *et al.*, 1991; Kszos *et al.*, 1992). An excellent example of use with a municipal effluent is provided by Ankley and Burkhard (1992).

A sample is subjected to a logical set of physical and chemical manipulations, and the strength of toxicity before and after each treatment is used to identify

Table 5 Selected Examples of Quantitative Structure-activity Relationships (QSARs) for Some Classes of Chemicals (from Suter, 1993)

Endpoint	Equation
Neutral Organics	
96-h LC50, freshwater fish	$\log LC50 = -0.94 \log K_{ow} + 0.94 [\log (0.000068\ K_{ow}) + 1] - 1.25$
14-d LC50, freshwater fish	$\log (1/LC50) = 0.871 \log K_{ow} - 4.87$
96-h LC50, marine fish	$\log (1/LC50) = 0.73 \log K_{ow} - 3.69$
48-h LC50, daphnids	$\log (1/LC50) = 0.91 \log K_{ow} - 4.72$
96-h LC50, mysids	$\log (1/LC50) = 1.25 \log K_{ow} - 4.83$
16-d LC50, daphnids	$\log (1/LC50) = 0.64 \log K_{ow} - 3.27$
16-d EC50, daphnids	$\log (1/EC50) = 0.72 \log K_{ow} - 3.05$
3-h EC50, green algae	$\log (1/EC50) = 8.865 - 1.0446 \log K_{ow}$
Aldehydes	
96-h LC50, freshwater fish	$\log LC50 = 0.4487 \log K_{ow} - 3.314$
Anilines	
14-d LC50, freshwater fish	$\log (1/LC50) = 0.988 \log K_{ow} - 4.02$
Carbamates	
48-h NOEC, sea urchin stages	$\log NOEC = 3.509 - 0.7174 \log K_{ow}$
Esters	
96-h LC50, freshwater fish	$\log LC50 = -0.535 \log K_{ow} - 2.75$
Esters, Phosphate	
96-h LC50, freshwater fish	$\log LC50 = 2.9305 - 0.5178 \log K_{ow}$
Phenols, Chlorinated	
96-h LC50, freshwater fish	$\log (1/LC50) = 0.75 \log K_{ow} - 3.51$ [at pH 6.1]
96-h LC50, freshwater fish	$\log (1/LC50) = 0.58 \log K_{ow} - 3.20$ [at pH 7.3]
96-h LC50, freshwater fish	$\log (1/LC50) = 0.46 \log K_{ow} - 3.04$ [at pH 7.8]
Phenols, Substituted	
96-h LC50, freshwater fish	$\log (1/LC50) = 0.73 \log K_{ow} - 0.61$ [at pH 6.0]
96-h LC50, freshwater fish	$\log (1/LC50) = 0.61 \log K_{ow} - 0.38$ [at pH 7.0]

toxic fractions and guide chemical analyses. Identification of toxicants is efficient because chemical analyses are done only on those fractions showing toxicity, and only for the types of chemicals likely to be found. Nevertheless, Canadian consulting companies caution that a complete TIE can sometimes be enormously time-consuming and expensive, and can yield equivocal results, or indeed, no identification of the toxicant(s).

A standard TIE consists of three phases. In Phase I *(Initial Characterization)*, aliquots are subjected to a sequence of simple manipulations (central downward path in Figure 6) that might include aeration, filtration, chelation of metals with EDTA, oxidant reduction or complexation with sodium thiosulphate, pH adjustment, or solid-phase extraction in which fractions are bound to an organic matrix and then eluted with methanol. After each manipulation, the toxicity of the aliquot or fraction is assayed with simple tests and compared with original toxicity. Each step in the procedure should unambiguously treat or remove a class of potential toxicants.

If the cause of toxicity were not clearly resolved in those Tier I manipulations of Phase I, then Tier II might repeat the separations on subsamples rendered acid (pH 3, left branch of Figure 6) or basic (pH 10, right branch), to change the state, solubility, or chemical reactivities of the constituents.

Phase I should have suggested the nature of the toxic components (e.g., metals or organics). Phase II *(Identification)*, applies the appropriate chemical measurements and toxicity tests, to attempt to identify the toxicants present. Phase III *(Confirmation)* determines whether the findings of the previous two phases can be substantiated. The third phase can use one or more techniques, such as toxic units analysis, or spiking by adding back appropriate concentrations of identified toxicants to see whether the original toxicity is recreated.

Sensitivity and use of a small volume of sample are important criteria in choosing test species and methods in TIEs. At least two species should be used for initial tests, because tolerance of different chemicals might vary (Kszos *et al.*, 1992). Any small-volume test is acceptable, but *Daphnia, Ceriodaphnia,* and fathead minnow larvae are often used. Simplified multiple-concentration tests or full-strength screening tests might be used in Phase I because the presence and relative strength of toxicity is more important than precision (Ankley and Burkhard, 1992). Definitive toxicity tests are used in Phase III.

A first step in conducting a TIE is to establish whether there is consistent strength of toxicity in the substance being studied. A variable effluent would probably require several evaluations. There are three other key elements in an effective TIE. First, there should be flexibility in adapting the general procedure to the nature of the sample [34]. Second, toxicity must be tested before and after each fractionation or chemical treatment. Third, wastes can contain several toxicants, so a major tactic is to remove the dominant one as completely as possible, to allow the remaining ones to show their effect and be identified.

Sublethal toxicity tests are now commonly used as part of the TIE, although the original methods used acute lethality. Such evaluations could use streamlined adaptations of acute sublethal tests (e.g., stripped versions of EC, 1992a,b,c,d). A TIE on municipal effluent was based on fertilization assays with sea urchins and sand dollars, a test which required only 40 minutes of exposure and incubation (Bailey *et al.*, 1995). The sublethal procedures of Phase I (USEPA, 1991d) are similar to lethal TIEs, and Phases II and III are the same as for lethality. Procedures for marine situations include both lethal and sublethal tests (USEPA, 1996).

---

[34] The kind and sequence of manipulation should be adjusted with experience or knowledge of the sample. If it were known to contain one very toxic component, an efficient first step would be to remove that component. Such a flexible approach was described by Jop *et al.* (1991) for an effluent from a chemical plant. Pressure filtration removed suspended solids, then the liquid was purged with nitrogen to strip volatile organics. Ammonia, the suspected principal toxicant, was removed by passing the purged fraction through either cation resin, activated carbon, or specific ammonia-binding clays. The ammonia-free fractions were still toxic, so were passed through XAD-resin to bind organics, which were subsequently eluted with acetone and chloroform. The toxic organic fraction was fractionated (acid, base, and neutral), and the most toxic neutral fraction was finally fractionated into 17 organic classes by means of high-pressure liquid chromatography. Chemical analyses of fractions showed which compounds were associated with toxicity, suggesting that ammonia and an organic compound [2-(4-orpholinyl)-benzothiazole] were responsible for observed toxicity.

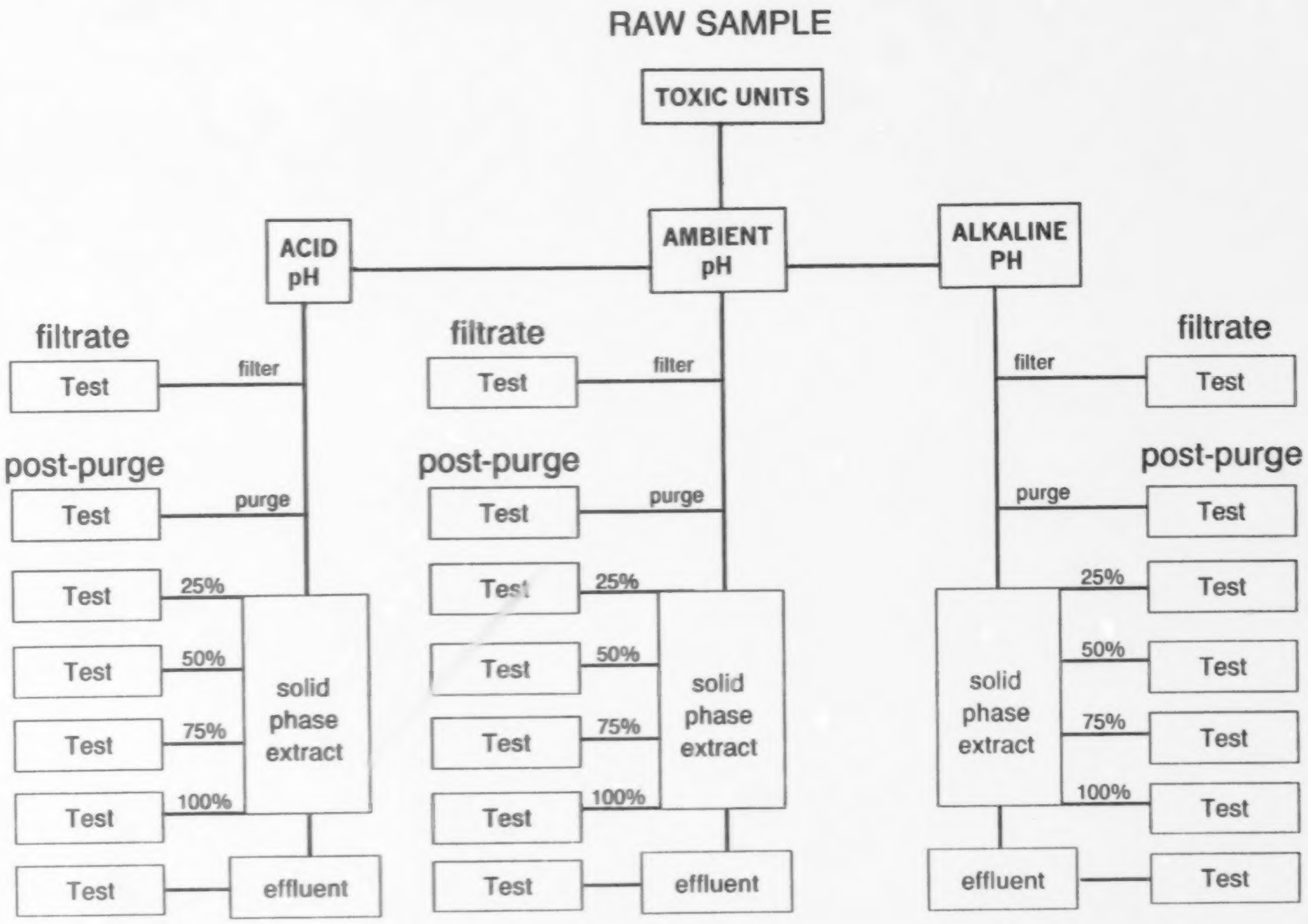
RAW SAMPLE
TOXIC UNITS
ACID pH
AMBIENT pH
ALKALINE PH
filter
purge
filtrate
post-purge
Test
25%
50%
75%
100%
solid phase extract
effluent

Figure 6 **Sequences of Manipulations of a Sample in Phase I of a Toxicity Identification Evaluation.** Characterization of the raw sample is in Toxic Units. Tier I of Phase I would follow the middle ("ambient") branch of the diagram. If that were not conclusive, Tier II manipulations could follow the acid and/or alkaline branches.

An approach using TIE would be efficient for assessing contaminated sites including soils and sediments. For sediment, there are tentative methods for TIE using acute lethality (USEPA and ACOE, 1994). One approach for sediments is to extract porewater and apply the methods for wastewaters (Ankley *et al.*, 1990). Another approach is to treat whole sediments by oxygenation or by adding chelators (alum, ferric chloride, lime) to reduce the bioavailability of some contaminants, then re-test for toxicity (Krantzberg and Boyd, 1992). Ankley and Schubauer-Berigan (1995) have reviewed the status of methods for TIE in studies of sediments.

Considerable research has been done for soils, but as yet, there are no accepted methods. Initial approaches test whole soil using earthworms or seed germination, then extract potential toxicants with water or solvents, re-test solid and liquid fractions, then proceed with further separations.

A Toxicity Reduction Evaluation (TRE) is more concerned with type of waste treatment for effective removal of toxicity. A TRE might be done after a TIE, or independently. Methods for municipal and industrial wastewaters are available (USEPA, 1988a, b).

4.8 Characterization of Effluents in Sweden

Key Guidance

- *A Swedish approach for characterizing industrial and municipal effluents uses a battery of integrated biological and chemical analyses to assess the potential environmental effects. Tests are designed to identify type of effect, bioaccumulation, and persistence.*
- *Three stages of testing are available, each stage representing a more detailed and sophisticated level of analysis, and based on the clues provided by results of the previous stage.*

Toxicity testing can be used to characterize a wastewater for its potential environmental effects, by constructing a toxicity profile in much the same way that a chemical profile is constructed. A Swedish scheme does this by using an integrated battery of biological and chemical analyses to assess industrial and municipal effluents (SNV, 1990).

In some respects the procedure is similar to the U.S. methods of toxicity identification described in the previous section, since one of its objectives is to identify the type of toxicant. In other respects the two procedures differ. The U.S. procedures are specifically intended for problem wastewaters, so that effective treatment can be designed (TRE). The Swedish program is more oriented to characterizing new or unknown effluents, to foresee and avoid problems.

The Swedish characterization measures biological effects, and might follow with chemical tests, to assess the degree of toxicity, resistance to biodegradation, and bioaccumulation. There are three stages of testing, and investigation proceeds to the next, more detailed level if results indicate the need for further study.

Stage I is initial screening of the wastewater, and uses a fixed battery of standard chemical and biological tests. Toxicity tests commonly include an LC50 for a fish, LC50 or EC50 for a crustacean, and a growth/reproduction test with an alga. Degradability is assessed by standard chemical or biochemical oxygen demand, an assessment that is useful for organic compounds, but obviously not for some substances such as metals. Similarly, bioaccumulation of organic substances is estimated by the lipophilic tendency of extracts, judged by thin-layer chromatography (TLC). Some standard chemical tests characterize the effluent in terms of common toxic components.

Stage II uses more sensitive tests, and is aimed at assessing effects after dilution. Sublethal toxicity tests could include growth and survival of fish larvae, reproduction of invertebrates and algae, physiological endpoints, or genotoxicity. One or two of these tests are chosen according to the nature of the toxicity observed in Stage I. Sophisticated chemical analysis proceeds according to the suspected problem(s).

Stage III seeks to verify cause-effect relationships between observed biological effects and identified chemical constituents. Possible techniques are varied and highly depend on previous results. The biological component could include morphological and physiological changes in caged or wild organisms. Bioaccumulation might be measured in fish from the site. Particularly for organic compounds, toxicological data are combined with information on degradability, bioaccumulation and chemical composition to decide whether discharge licences should be granted or renewed, and what restrictions should be placed on those licences (SNV, 1990). The whole exercise is intended to be a logical progression of effort and accumulation of relevant information, using good principles of scientific investigation.

4.9 Troublesome Sets of Results

Key Guidance

- *Unusual data for toxicity tests can stem from one of more of the following causes: stimulation at low concentrations; shallow, stepped, or inverted concentration-effect distributions; high variance; low variance within replicates; improper range of treatment levels; non-constant concentrations; missing data; and operator error.*

- *The following steps are recommended for systematically examining the results of a test, particularly a test with inconsistent data. (1) Ensure that the report is complete. (2) Review the relationship of effect to concentration. (3) Review the supporting test information. (4) Review information on quality control. (5) Assess reasonableness of findings for the tested substance and conditions.*

Toxicity tests use living organisms or living systems which have some inherent variability, and accordingly, test results might occasionally differ from typical concentration-response curves, even if strict procedures of quality control were followed. The documents describing methods usually specify the criteria for valid tests, and the precision required (see Sections 3.2, 3.3), but unusual results can arise in tests that meet the criteria. This section gives examples of unusual results, and provides some general guidelines for dealing with them. Similar problems were evident in a survey of 3000 toxicity tests on liquid effluents (WSDOE, 1998), for approaches that are used in the USA and Washington State (WSDOE, 1997).

Detailed advice on statistical analysis of irregular data is not given herein, but an investigator might consult statisticians familiar with toxicology, colleagues, appropriate statistical texts such as Newman (1995), or an Environment Canada guidance document (EC, in prep, a).

4.9.1 Potential Anomalies in Data-sets

This section considers sublethal effects; variations in toxicity curves for lethal tests are described in Sections 4.9.1 and 5.6.3. Four possible aberrations of sublethal tests are shown in Figure 7, which might represent inhibition of growth compared to the control, or another similar effect. The experiment might be intended to estimate the IC25. The solid straight line in each panel represents idealized results with little variation.

Stimulation or enhancement *(hormesis)*. The test substance has a slight stimulatory effect at low concentrations (Figure 7A). This is not unusual in sublethal tests, and the causes are seldom clear (reviewed in Stebbing, 1982). There is no entirely satisfactory solution for analysis. Simply fitting a line to estimate the IC25 could obviously lead to a value that was higher than if no hormesis had been present. Estimating the IC25 by the usual commercial computer programs would yield a lower value (i.e., the sample would be considered more toxic), because the programs adjust the control response to a higher value. An estimate of the NOEC (lack of significant difference from the control) could be influenced upwards by the interplay of stimulation and depression of response (e.g., growth). There are no standard methods for correction of hormesis, and the usual way of treating such results is to proceed with the unmodified method of analysis. It might be feasible to omit the anomalous lower concentrations (i.e., the stimulatory ones). In a case of strong hormesis at several low concentrations, it might not be feasible to analyze results to obtain the toxicological endpoint.

A shallow relationship of effect to concentration. There is no fundamental irregularity in the relationship except that a very large change of concentration is required to elicit an appreciable change in the effect (Figure 7B). Such a finding would be unusual. Nothing prevents the application of regular methods of analysis for the example shown in panel B, but any report should call attention to the unusual dose-effect proportions, and the questionable value of the calculated endpoint.

Such a test might not have a low enough test concentration, so that the effects were different from the control, at all concentrations. The NOEC could not be calculated in that case, although the IC25 could be, if it were bracketed by test concentrations.

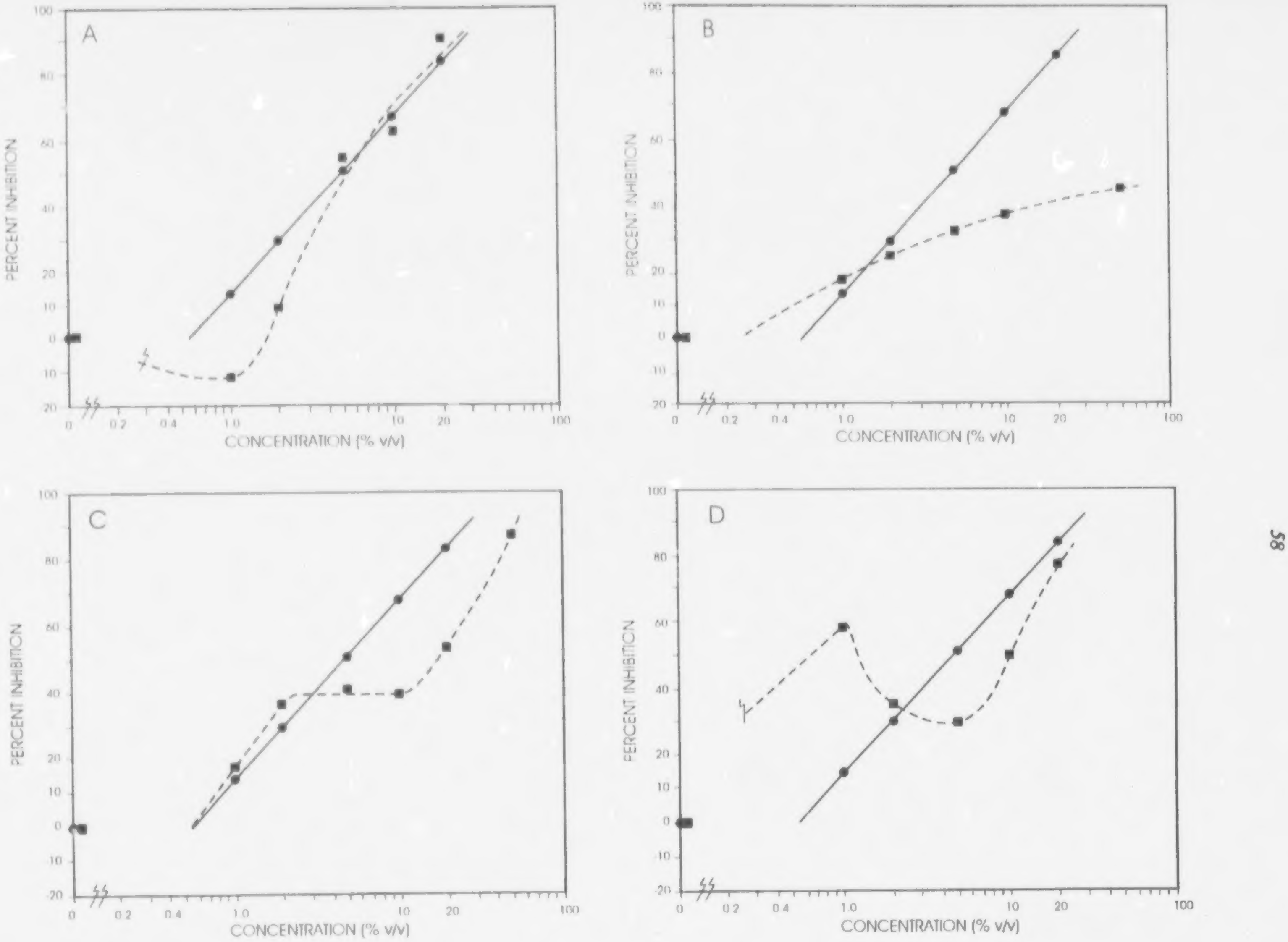

Figure 7 **Unusual Concentration-effect Curves.** An idealized straight-line response is shown by the solid line and its circular points, representing percent inhibition of some sublethal response such as growth. A: Stimulation at low concentrations. B: A shallow effect-curve. C: A biphasic effect indicating two modes of action. D: Inverted effect-curve.

Biphasic or stepped effects. The hypothetical results in Figure 7C suggest that there are two modes or mechanisms of toxic effect. One shows a normal dose-effect pattern at lower concentrations, but about half the organisms appear to be unaffected by this mode of action, so there is no increase in effect over a moderate middle range of concentrations. At higher concentrations, a second mode of action comes into play, initiating a second dose-effect relationship that appears normal in those upper ranges of concentration. This would be quite an unusual finding but it occurs, and attention should be called to it in any report. It might be possible to provide estimates of toxicity for both modes of action, but the lower range of concentrations would be the important one in sublethal testing.

Inverted relationship of effect to concentration. Toxicity decreases from the lowest concentration into a middle range, then increases in a more or less normal fashion (Figure 7D). Rarely, this might occur by chance if individual test organisms were extremely variable in their response, and there were few of them at each concentration.

A more likely cause of such an inverted pattern, however, would be gradients of ancillary chemical conditions over the range of concentrations, resulting from differences in the basic chemical nature of the sample being tested and the dilution water. For example, pH has an effect on toxicity of many chemicals, and a particularly strong effect on toxicity of total ammonia to fish. Ionized ammonium (NH_4^+) dissociates to unionized ammonia (NH_3) at basic pH. The un-ionized ammonia is very toxic to fish but the ammonium ion has little toxicity. The example in Figure 7D might be an effluent that had a high concentration of total ammonia, much of it present as ammonium ion because the effluent was slightly acidic. A well-buffered, alkaline laboratory dilution water could influence pH in such a way as to increase the proportion of un-ionized ammonia at low concentrations, enough to cause greater toxicity. Another example encountered in Environment Canada laboratories involves mine-water effluents which are very hard because they have been treated with calcium carbonate. If the test dilutions are made with a very soft laboratory water, an inverted curve can be obtained because the low concentrations are soft water, making the metals from the effluent very toxic.

High variance. High variance in test results does not necessarily represent an anomaly, but it might cause problems in interpretation. If the coefficient of variation were greater than the 20–30% that might be expected, the resolution of an endpoint such as the NOEC would be obscured. An overly high estimate of NOEC is, in fact, a major problem that occurs with variable tests, and the endpoint might be at odds with evidence from a linear change of effects with concentration. Use of ICp or ECp as an endpoint might be feasible and more representative of the data.

Low variance within replicates. Low variance is a less common problem, but might result from very uniform test organisms, proficiency in running the test, or partly from chance alone. The variance in controls and within each treatment level might be much smaller than a normal 20%. If the NOEC were being estimated by analysis of variance and multiple-range tests, certain treatments (concentrations) might be significantly different from the control, in an unusual or unexpected pattern. The best solution might be a regression-type analysis such as ICp.

Unsuitable range of concentrations. This, and the following three difficulties, are not anomalies of the data, but are difficulties usually caused by practices in the laboratory. The concentrations selected should, at the end of the test, bracket the endpoints of interest. In many sublethal tests, concentrations above and below the endpoint are required, otherwise the endpoint cannot be calculated and/or the test is not valid. It is a common problem to have an unsuitable range in the test. This is often caused by incorrect anticipation of the endpoint's value, before the test is run, and by selection of a range of concentrations that is too narrow and/or too high or low. The obvious remedy is to conduct a preliminary range-finding test. Another remedy is to use a relatively wide range of concentrations. The resulting spread between adjacent concentrations might increase the uncertainty associated with the endpoint, but at least there would be an estimate rather than a report stating "greater than x%" or "less than y%".

Changing concentrations. Toxic components can volatilize, decompose, sorb to test apparatus, or cease to remain stable for a number of reasons. The end-result is concentrations that change during the test. This might happen and not be recognized [35]. A drop in concentrations (or a rise) could produce nonconforming data or unusual distributions similar to those shown in Figure 7. The problem can be minimized by following all good practices of the test method and using better than minimum requirements. Knowledge of the nature of the test substance can assist greatly in avoiding these problems (and in interpreting the results).

Missing data. A replicate might be missing because of a handling error, or an accidental loss of data for one or more replicates might happen even in valid tests. Short of repeating the experiment, the only solution is to turn to methods of analysis that do not require the missing information. Methods that exploited trends, such as regression and ICp, would likely be useful.

Operator error. Unusual results might sometimes originate, not with the test organisms or conditions, but with people who collect the information. This would commonly include errors of transcription or entering values into a computer program. "Data entry and arithmetic errors are the most common mistakes currently being made by labs ..." in reports to the state of Washington (WSDOE, 1998). Other errors could be improper use of statistical methods, or simply a lack of training and experience in recognizing problems and dealing with data. Differences of interpretation by individuals could also give rise to inconsistencies. The remedy is to set up and follow good practices of quality assurance and quality control.

Problems with controls. Unacceptable responses in the control are a common difficulty. Such responses signify a problem with the test and cannot be ignored or circumvented. In most test methods such as those of Environment Canada, the test is considered invalid if the criteria for the control are not met, and the only acceptable follow-up action is testing a new sample or a new preparation of a test substance.

Sometimes, the data might be used with caution to provide an approximate result suitable for general orientation. Some quantal endpoints (LC50, EC50) do not involve the control in direct calculations, so there is no impediment to calculating the endpoint by the normal methods. Nevertheless, whatever factor caused an effect in the control was probably also operating in the test concentrations, confounding the result for effect of the toxicant. There is no entirely satisfactory method of correcting for the control effect in these quantal tests (EC, in prep., a). If the control effects were accompanied by an otherwise normal dose-effect curve, in other kinds of tests which incorporate the control response into calculations of endpoint, the data from the lowest treatment level might be used with discretion as a substitute for the control. In some tests, it might be appropriate to combine data from the lower treatment levels for statistical comparisons against the higher concentrations. Any of these tactics would have to be supported by the particular pattern of results, and/or parallel information on performance of controls in other tests. It must be emphasized that such stratagems to provide an approximate endpoint of toxicity would only be useful for general guidance.

4.9.2 Dealing with Problems in Test Results

The following five steps are recommended as a general framework for checking sets of results, and detecting unusual ones. The approach applies to valid and invalid tests, but this discussion assumes validity unless otherwise noted.

1. **Ensure that the report is complete.** Information required in a report is specified in Section 8 or 9 of each biological test method published by Environment Canada, and the same type of information might reasonably be expected in reports for nonstandard and novel methods. Comments should be included in the report for any deviations from accepted practice and any unusual data.

2. **Review the concentration-effect data.** The basic original observations should be available in any report and should be inspected. The easiest way to check the relationship of concentration and effect is graphically; if a graph is not included in the report, one should be plotted [36]. Anomalies such as those in Figure 7 will be evident, as

---

[35] Most methods for testing toxicity of wastewaters and contaminated sediments or soils do not require measured concentrations of toxicant(s). Indeed, there might not be anything representative that can be measured, or it might not be known what to measure. Hence, there is often no evidence to demonstrate the changing concentration of toxicant during a test. That absence of evidence does not mean that toxicant concentration remained stable.

[36] Errors in transcribing data and entering them into a computer program can occur (WSDOE, 1998). Accordingly,

will any deviation of computed endpoints from the distribution of data.

3. **Review the supporting measurements.** Physicochemical and other background data should be scrutinized with special care if problems were evident in the dose-response data. The purpose is to note anything unusual about the test procedures or possible errors. Some test-method documents specify the observations that must be reported (e.g., most Environment Canada methods require notes on unusual behaviour of test organisms, or changes in the appearance of test solutions). Certainly, any deviations from the method must be reported. Physical and chemical measurements could be particularly useful for interpreting deviant results. There should be consistent patterns in physicochemical measurements on different days and treatments.

4. **Review quality assurance information.** (see Section 3). Environment Canada test methods specify certain information. Items concerning holding or culturing conditions for the test organisms and use of reference toxicants are particularly important.

5. **Assess reasonableness of results for the substance and conditions.** The final step is to evaluate the results in light of the information uncovered on the nature of the test substance and the procedures used. That could include a range of things such as previous results on the substance, results for other tests conducted at the same time on the same substances or additional ones, or other tests with the same stock of organisms.

These five steps can indicate ways of interpreting and increasing the value of unusual data, perhaps simply by correcting errors or by alternate methods of analysis. Reviewing the data might indicate that results were not dependable, in which case there should be no hesitation in rejecting them.

4.10 Toxicity of Effluents

Toxicity tests on effluents are commonly used in Canada and elsewhere. In this document, an effluent is any wastewater from domestic, municipal, or industrial sources, discharged to a surface waterbody through a pipe or channel. In many cases, the quality of effluent is controlled federally through regulations under the *Fisheries Act*. However, provincial legislation or policies also use toxicity to regulate effluents. A book edited by Grothe *et al.* (1996) provides a multi-facetted review of toxicity tests on effluents, particularly for the approaches used in the USA.

Methods for aquatic toxicity testing were originally developed for assessing effluents, but have now been adapted to fluids from virtually any source, including sediment porewater and extracts from soil or other solid material. "Wastewater" is used in this section to include effluents, leachates, and elutriates; the latter two fluids are more specifically covered in Section 4.11.2.

Toxicity tests on wastewaters provide an orderly basis for assessing, comparing, and regulating waste materials, even if the specific toxicants are unknown. The tests integrate the total effect of individual chemical components, and as in other situations, complement chemical analyses in detecting problems and remedying them. The mixing, dilution, and dissipation of wastewater toxicity can be modelled or measured in the receiving water the same as any single chemical component (see Section 4.10.2).

Effluents tend to vary in composition an degree of toxicity. Most effluent regulations specify the frequency of sampling (e.g., weekly, monthly, quarterly), the procedure for collection, and sometimes the sample volume. If conditions are not specified, the sampling design should have a goal of obtaining a representative measure of the potential effects of the effluent, considering variation in its quantity and quality. A grab sample (collected more or less instantaneously) provides a snapshot of effluent quality, and can be adequate for preliminary assessment, for compliance audits, spot-checking during a spill, etc. Composite samples represent a series of instantaneous samples pooled over time, and smooth out variation. Composites over 24 h are often used for industrial and municipal effluents, with shorter periods for storm drainage and other sources that respond to predictable events.

---

one should beware of using a computer program to plot from a file of data that was created at the time the report was originally compiled. Errors in entering data would be perpetuated, and incorrect estimates of endpoints would not be detected. The original observations should be available for checking. For electronic reporting, the computer file might be the only listing of raw data, in which case it should be inspected for reasonableness and plotted.

The test species do not necessarily have to be resident in the waterbody receiving the effluent. Standard species produce comparable data from place to place, and are chosen to be sensitive surrogates for indigenous communities (see Section 4.1).

4.10.1 Monitoring Effluents

Key Guidance

- *The rationale for monitoring an effluent with single-species toxicity tests should be clearly identified and periodically re-evaluated to ensure that the species tested, monitoring frequency, and pass/fail limits remain appropriate.*

- *Failure of a toxicity test should be followed up quickly with a second test to confirm or challenge the first result. If results differ, compare chemical profiles of the two samples and examine data for quality control, operations at the plant, and other records.*

The toxicity of effluents might be monitored for one of several reasons: (1) to assess compliance with discharge limits; (2) to predict the site-specific impact in the receiving water, from this discharge or a combination of discharges; (3) to quantify the combined toxic effect of diverse components that have been combined into the effluent; and (4) to assess improvements with changes in industrial process or technology for pollution abatement. The rationale for monitoring should be clearly identified and periodically re-evaluated to ensure appropriate strategies and methods.

In checking an effluent for regulatory compliance, most prudent environmental managers would quickly follow up a failed toxicity test with a second test. Ensuing information would indicate whether the problem was consistent or due to a temporary mischance or error. Some pulp mills in Alberta routinely collect duplicate samples and send one to each of two testing laboratories, if results are intended to fulfil regulatory requirements. Consecutive or duplicate tests might disagree for one of several reasons: an atypical pulse of toxic substance; temporary interruption or short-circuiting of the waste treatment system; or problems in collecting samples, or handling, storing, or testing them. A brief episode of toxicity might be traced to particular operating conditions if chemical analyses were available for toxic and non-toxic periods, or if samples had been stored.

The records for failing and passing toxicity tests should be examined for information on quality assurance and quality control, statistical analysis of results, and other relevant items. Difference between laboratories might be due to the modifying action of different dilution waters (Section 4.15), or different sources of test organisms. Results with reference toxicants might reveal differences arising from those causes. Warning charts should help indicate whether particular test runs were within normal standards of performance.

If a sample fails or almost fails a compliance test, remedial action would be appropriate. For an industrial facility, that might include immediate reporting to the government regulatory body, an in-plant audit of processes and sewers, temporary diversion of effluent to a collection facility, or initiation of procedures to identify the toxicants (Section 4.7).

4.10.2 Predicting Water Quality and Regulating Toxic Effects

Key Guidance

- *Water quality guidelines are used to judge the quality of surface waters. The guidelines can be chemical concentrations or statements on degree of toxicity.*

- *The USEPA developed a specific and detailed procedure for limits of receiving-water quality, based on lethal and sublethal tests with effluent and specific toxic substances. Toxicity testing can be used to screen effluents, set discharge limits, and enforce regulatory compliance in a consistent manner.*

- *The EPA suggested a surface-water criterion of 0.3 Lethal Toxic Units, not to be exceeded for more than one day, nor more frequently than once every three years. A second suggested criterion was 1.0 Sublethal Toxic Units, not to be exceeded for more than four days, or more often than every three years.*

- *The EPA estimated in-stream concentrations of toxic substances by a simple mass-balance equation. Regulatory compliance required that toxicity criteria be met at the extreme 7-d average low flow expected once a decade. Variability in toxicity should be encompassed by calculating on the basis of the 99th percentile for Sublethal Toxic Units, as expected from measured toxicities. The calculations were apparently to use the median daily flow of effluent. The influence of any sublethal toxicity in upstream waters was not accounted for by this approach.*
- *The EPA did not specify the size of mixing zones since that is the prerogative of individual states in the USA. In Canada, Quebec, Alberta, and British Columbia have a specified maximum size for mixing zones.*
- *Guidelines can provide information for deriving regulatory limits in surface waters. These might be maximum chemical concentrations, or a fraction of a sublethal threshold.*
- *Instead of monitoring toxicity in samples of receiving water, the effluent can be tested, with calculations to the expected effect in the waterbody. Monitoring the effluent can have advantages of convenience, speed, and accuracy.*
- *Regulations proposed for Australian pulp mills included chemical and biological limits, and no sublethal toxicity beyond a 500-metre mixing zone.*

Chemical and physical guidelines for judging the quality of receiving water and sediments are well established (e.g., CCREM, 1987; OMOE, 1991). Guidelines which are listed for aquatic life can be used to predict the suitability of a waterbody for supporting a biological community. If the concentrations measured in the environment exceeded the guideline, deleterious effects on the community would be predicted.

In a similar way, toxicity tests can be used to judge the quality of receiving water. That can be done by directly testing samples of the surface water (so-called "ambient" tests). Examples would be testing the toxicity of a sample of sediment from the bottom of a harbour, or holding a cage of fish in a river (that would also be an *in-situ* test). A quality guideline for toxicity would usually be a simple and straightforward one, such as "no observed sublethal effect on growth or reproduction" in a specified test.

Using chemical or toxicity guidelines is a major tactic for evaluating or regulating environmental pollution [37]. When dealing with discharges of effluent, however, it is not necessary to apply quality guidelines by actually sampling and measuring the receiving environment itself. The guidelines can be applied by making measurements on the effluent, then calculating the conditions expected in the environment, by accounting for dilution. Measurements on the effluent are generally speedier, more convenient, and more definitive since any toxicants are at their highest concentrations. It would be desirable, of course, to periodically confirm the predicted effects by testing the receiving water or by field surveys.

Regulations for pollution control can be derived from both chemical and biological quality guidelines. For example, a regulation might require that "average concentration of copper in the receiving water should not exceed x.x μg/L, and maximum concentration at any time or place should not exceed y.y μg/L". In similar fashion, a regulation might state that "lethal toxicity should be absent in the immediate discharge zone, and sublethal toxicity should be absent beyond the limits of the mixing zone". The test organism, method, and size of mixing zone would need to be specified. Some recommendations for designing an effluent monitoring program come from a carefully planned program for Washington state (WSDOE, 1997), followed by a detailed examination of test results and observed ecological damage in waterways (WSDOE, 1997). One recommendation for increased effectiveness was to use simple, rapid, and economical toxicity tests on the effluent, and to do them more frequently.

---

[37] As described in Sections 1.2 and 4.4, strategy for monitoring or regulating pollution can be broadly divided into three tactics: (1) effluent limits, an approach based on Best Available Technology; (2) quality guidelines applied to the receiving environment, which is the topic being covered here; and (3) biological surveys as a definitive audit of whether satisfactory conditions prevail.

In this monitoring or regulatory scenario, the toxicity testing is relatively straightforward. If concerned with an effluent, its sublethal toxicity could be measured using a standard test such as reproduction of daphnids, leading to an estimate of the IC25. If expected concentrations after dilution in the environment were greater than the IC25, then obviously, sublethal effects would be predicted in the environment. The difficult part might be obtaining an accurate estimate of the available dilution, and the pattern of mixing. Various models might be used. In a river, for example, available dilution is often based on the lowest average flows expected once in 10 years (see the following section). Mixing zones would generally have to be established by field surveys, although computer models can assist (some are listed in USEPA, 1991a). The state of Washington suggests minimizing such difficulties of predicting dilution by moving many or all of the toxicity tests from the effluent to the receiving water, the so-called "ambient tests" (WSDOE, 1998). However, direct testing of the effluent has very strong advantages of (a) exact knowledge of the waste concentrations, and (b) maximum toxic effect so that one would not overlook a toxic contribution that just failed to elicit a test response because of dilution.

An example of straightforward regulatory limits for receiving-water quality is the suite of limits proposed for Australian bleached kraft pulp mills (Australia, 1989). Limits were listed for water and/or sediments, at the edge of marine mixing zones set at a diameter of 500 m. Physicochemical limits included those for chlorates, categories of organochlorines (AOX, EOX), chlorinated phenolics, dissolved oxygen, suspended solids, and colour. Absence of sublethal toxicity had to be demonstrated at the edge of the mixing zone for a fish, invertebrate, and burrowing organism; the aquatic tests were to be done on the effluent and calculations made to dilutions outside the mixing zone. Another example is the more formal and complex procedure developed in the USA for regulating toxic discharges, a procedure reviewed in the following section.

Regulating Toxic Quality of Water in the USA. The USEPA (1991a) developed a very formal use of toxicity, to regulate the discharge of effluents. This approach was called "Water Quality-based Toxics *(sic)* Control", and was formulated under the United States Clean Water Act. The EPA uses toxicity testing to set discharge limits and enforce compliance in a consistent way that still accounts for site-specific conditions. Water quality is defined both in numerical chemical standards and in narrative toxicological standards, and both sets of parameters are considered identically. The broad toxicity approach includes:

1. limits for acute lethal toxicity in the receiving water;
2. limits for sublethal exposures in the receiving water;
3. restrictions on frequency and duration of excursions beyond the limits; and
4. procedures for dealing with variation in quality of effluent.

The first step in the USEPA approach is to convert all toxicity measurements into lethal or sublethal Toxic Units (LTU or SLTU, see Section 4.3.1). Results are thus expressed in units similar to chemical concentrations. An effluent can be defined in terms of toxic loading by multiplying the rate of discharge by LTU or SLTU (Section 4.3.3). In the case of an effluent discharging to a river, estimates of in-stream concentrations can then be made using a simple mass balance calculation:

$$C_d = [(C_u \times Q_u) + (C_e \times Q_e)] / Q_d$$

where: C_d = toxic concentration downstream (TU)
Q_d = downstream flow (m^3/s)
C_u = toxic concentration upstream (TU)
Q_u = upstream flow (m^3/s)
C_e = toxic concentration in effluent (TU)
Q_e = effluent flow (m^3/s)

Calculated in-stream concentrations can then be compared with the suggested criteria of 1.0 SLTU and 0.3 LTU [38]. The value of 0.3 LTU is intended to eliminate acute lethal effects on virtually all individual organisms. The fractional 0.3 LTU cannot be measured in the receiving water, as such, being analogous to a chemical criterion that is below analytical detection limits. It can be reasonably handled, however, since it results from calculations of dilution from the much stronger concentration, measured in the effluent.

---

[38] In the U.S. technical support document (USEPA, 1991a) the sublethal toxic units are designated as "chronic" TU and those for lethal effects are designated as "acute" TU.

The USEPA (1991a) suggests that if the lethal criterion (0.3 LTU) is not exceeded for longer than one day, and no more frequently than once every three years, then aquatic life should be adequately protected. The suggestion is based on observed recovery of disturbed ecosystems described in Section 4.10.3. Similarly, for the sublethal criterion of 1.0 SLTU, a maximum duration of four days with a return frequency no shorter than three years is suggested.

The EPA approach is illustrated in the following, using a riverine system which is a simple one to consider because of the unidirectional flow of water. The same principles apply to lakes or marine systems, but dilution is complicated and might need to be measured directly; "upstream" water would be any part of the waterbody in which effluent concentrations were not elevated above background (details in USEPA, 1991a).

For a river, the simplest approach uses a steady-state, "design" value for each of the variables in the previous equation. Low river flows are used since toxic conditions would be worst then. One option is to use the flows known as 1Q10 and 7Q10. The 1Q10 is the one-day low flow that is expected as a minimum which would occur only once in 10 years, and 7Q10 is the 7-d average low flow that is expected only once in 10 years. If the toxicity *just* met the requirements ($\leq$0.3 LTU or $\leq$1.0 SLTU) at the 7Q10, it happens that there would be approximately one transgression of the limit every three years, (i.e., just at the greatest frequency suggested by the EPA)[39].

The variable toxicity of most effluents must be considered in steady-state calculations with the equation. A worst-case value is used for toxicity. The 99th percentile value of LTU and SLTU is calculated, using existing data from toxicity tests. The EPA gives step-by-step instructions for completing that calculation.

Two variables are not well described in the Technical Support Document of USEPA (1991a). The first is the variation in effluent flow rates. The guidance document implies that median values should be used but provides little insight on the rationale. The USEPA Region VIII (1994) recommends using mean daily effluent flow rates when evaluating sublethal toxicity and maximum daily flow rates for lethality.

The second poorly defined variable is the background river concentration, an important item that is almost impossible to measure directly. Toxicity from upstream waters might exist in a multiple-discharge situation, especially with overlapping effluent plumes. If upstream water were thought to be contributing to downstream toxicity, effluent tests might use the upstream water as the dilution water. That would account for any toxic contribution and also the action of any factors modifying toxicity (Section 4.15). If upstream water was used for dilution, the toxicity results for a given sample could be used directly in calculations, by setting the background toxicity (C_u in the equation) to zero, because the effect of any upstream toxicity would be already contained in the measurement of effluent toxicity. Diluting tests on a downstream discharge with toxic upstream river water is not entirely satisfactory, since the total toxicity is ascribed to the downstream discharger, and none to the upstream discharger. If the upstream water is overtly toxic, separate tests could be made on it and the effluent, with calculation and comparison of separate toxic units.

The size of mixing zones, within which the toxicity criteria would be exceeded, is not specified by USEPA (1991a) because mixing zones are to be decided by individual states. EPA Region VIII (1994) published a default policy for mixing zones: (1) lethality criteria must be met at the end of a discharge pipe; (2) sublethal criteria must be met after dilution with no more than 10% of design low flow for a river, or 4:1 dilution in lakes or reservoirs.

Canadian Regulations of Toxic Water. In Canada, some use has been made of limitations of sublethal toxicity in receiving waters, but without the formal complexity of the USEPA method. Regulations under the Fisheries Act stipulate no net loss of fish or fish habitat, and there is no provision for a mixing zone in which sublethal damage would be allowed. Mapping of plumes is required under the Canadian program of environmental effects monitoring for the pulp and paper industry (see Section 5.2.4). The values mapped are 1% effluent concentration or 0.1 of the IC25, much lower than those mentioned from USEPA Region VIII (1994). British Columbia defined a mixing zone of 100 m for pulp mills, describing it as an *initial dilution zone* in which water quality guidelines might be exceeded (Pollution Control Board, 1977). Quebec has used specific definitions of allowable mixing zones since 1980, and these are described under "Mixing Zones", Section 5.2.7. Alberta has implemented an approach which draws some of its elements from the USEPA system, but is more flexible (AEP, 1995). Alberta recognizes a small zone (plume) of 30 m which might exceed specified *acute instream guidelines* (i.e., specified chemical concentrations). In a surrounding zone, *chronic instream guidelines* might be exceeded, i.e., aquatic organisms might be sublethally affected in exposures longer than four days. A number of restrictions also apply, to protect other uses, and the waterbody as a whole (see Section 5.2.7).

---

[39] A more realistic, but complicated technique uses a continuous set of data to define each of the variables in the equation, then calculates a continuous set of downstream river concentrations. This can be done using a time-series simulation, or using a computer to permute the real values to calculate a frequency distribution of concentration. The resulting time-series or frequency distribution of downstream effluent concentrations can be compared with the allowable frequency of transgressions. These more complex analyses might be used if the steady-state approach did not yield firm conclusions.

4.10.3 Recovery of Damaged Aquatic Ecosystems

Key Guidance

- *The long-term health of aquatic ecosystems might be maintained despite occasional severe surges of pollution, if sufficient time had elapsed between events.*

- *At least half of damaged rivers show appreciable recovery of species and numbers of fish in two years, while invertebrates show better recovery in that time, essentially complete in 70% of cases studied. Longer recovery times would be needed if there were severe or multiple disturbances, persistent bioaccumulative toxicants, a long-lived species, or if there were no refugia for recolonization.*

If an unusual toxic discharge were to cause severe damage in the receiving water, the rate of recovery would depend on the type and severity of the disturbance, its duration, the lengths of life-cycles of the species affected, and the potential for recolonization. Some analyses are available for aquatic communities (fresh water), but apparently not for other kinds of communities.

Fish populations in most river systems show some degree of recovery from short-term major disturbances in two years, according to a review of 49 cases which resulted mostly from toxic chemicals (USEPA, 1991a). Half of the species of fish re-appeared in 67% of the systems and populations recovered to half of the original in 52% of the systems. Benthic communities showed somewhat faster recovery. In two years, most invertebrate communities recovered to former levels of species richness (71% of communities), total density (75%), and total biomass (71%) (Detenbeck *et al.*, 1992).

Some studies of chemical spills have found recovery in populations of aquatic invertebrates in a year or less (Heckman, 1983; Van Urk *et al.*, 1993). A stream that was dosed with pesticide showed 90% reduction in the population density of invertebrates, then recovery was rapid in the first year and complete in two years (Wallace *et al.*, 1991). Processing of leaf litter, an essential ecological process mediated by the invertebrates, was also projected to reach pretreatment levels in two years.

In the preceding examples, recovery was appreciable in two years, but obviously not complete in most cases. Periods longer than two years for substantial recovery would be required for unusually severe disturbances, multiple disturbances in close succession, persistent bioaccumulative toxicants, or recovery of a long-lived species of fish (Barnthouse *et al.*, 1990; USEPA, 1991a). Recovery can be slow if it must depend on reproduction following chance recolonization, or if dams or other barriers prevent immigration (Detenbeck *et al.*, 1992). Given areas of refuge, or unaffected upstream or downstream areas, recovery through recolonization can occur within one year.

Other studies indicate similar recovery rates. A small stream disturbed by a severe chlorine spill took <1 year for recolonization by algae, invertebrates, and non-territorial fishes (Heckman, 1983). Northern pike suffered a year-class failure, however, and were expected to need at least two or three years, being a resident territorial species. Severe disturbances of salmonid populations can require >5 years for recovery if recolonization does not occur, because repopulation does not begin until surviving young reach sexual maturity (Detenbeck *et al.*, 1992; Curtis, 1990). After a major pesticide spill in the Rhine River, recovery of benthic invertebrates depended on the generation time of a species. Chironomid midges, with many generations per year, recovered in a few months, while some caddisflies required six months, and a particular species of caddisfly needed a full year (Van Urk *et al.*, 1993).

The type of community has a strong influence on recovery time. Headwater streams are unstable and undergo periodic natural disturbance, so they tend to support biota that are adapted to rapid recovery. More constant systems, such as large lakes and slow rivers, tend to support larger, longer-lived species that are less tolerant of disturbance and slower to recover afterward (Detenbeck *et al.*, 1992). Paradoxically then, frequently disturbed systems could have greater capacity to recover from toxic discharges (Van Urk *et al.*, 1993).

The type of pollutant also has a strong bearing on recovery time. Clearly if a persistent, bioaccumulative toxicant (e.g., mercury, cadmium, persistent insecticide) were heavily dosed into an ecosystem, it could take many years or decades for full recovery. Type of waterbody and turnover times for water would play strong roles in determining the time for cleaning. Under Canadian boreal conditions, the toxic metabolites resulting from spraying the insecticide DDT took a full decade to work themselves out of a salmon river and approach concentrations in a clean system (Sprague *et al.*, 1971). Effects on returning adult salmon seemed discernible for more than half a decade (Elson, 1966; 1974).

The USEPA (1991a) recommends that toxicity-based effluent limits should not be exceeded more than once every three years, on average, if receiving environments are to be "adequately" (not perfectly) protected. That rate might be a reasonable goal in some cases, judging by research to date, but longer intervals between transgressions would be required for specific sites and particular kinds of ecosystems.

4.11 Toxicity of Sediment

Many aquatic toxicants are hydrophobic and tend to sorb to the surfaces of substances in sediment, or to suspended solids which eventually settle into the sediments. Benthic sediments tend to be a sink which accumulates metals and insoluble organic and inorganic compounds (Giesy *et al.*, 1990; Alden, 1992). Some sediments in rivers, lakes, estuaries, and nearshore marine environments near cities or industrial facilities contain high levels of contamination (e.g., Krantzberg and Boyd, 1992; Tay *et al.*, 1992). Such sediments can serve as a *source* of contamination for the water column and its organisms, particularly if the sediment is disturbed (Chapman, 1989b; Giesy and Hoke, 1989).

More directly, toxicants in sediment can pose a hazard to benthic organisms such as insects, worms, molluscs, and bottom-dwelling fish, and damage to them could have major effects on functioning of an ecosystem. In some cases, toxicants in sediments could be much more important than concentrations in the water. Streams in Colorado exceeded water quality criteria for selenium consistently, without apparent effects on the biota, and it appeared that chronic selenium toxicity was governed by concentrations in the sediment, and dietary uptake through movement into the food chain (Canton and Van Derveer, 1997).

Sediments tend to vary in structure and composition, even around one site. Physicochemical differences in characteristics such as particle shape and size, or organic content can profoundly influence bioavailability of toxicants and the degree of toxicity. Those differences must be considered in interpreting toxicity tests with sediments, and appropriate controls can be particularly difficult to set up (Luoma and Carter, 1993).

The rationale for choosing appropriate methods and organisms for toxicity tests with sediment has been reviewed by Giesy and Hoke (1989) and Lamberson *et al.* (1992). Burton (1992) reviewed test methods and field evaluation. The technical aspects of assessing the ecological risk of metals in sediments is discussed in a series of 15 papers, starting with an overview by Ankley *et al.* (1996). A critical evaluation of the whole field of testing the toxicity of sediments, particularly with regard to risk assessment, has been provided as the proceedings of a workshop (Ingersoll *et al.*, 1997). Background on the interaction between water and the sediments and contaminants contained in the sediments is provided in a compilation of 72 papers which emphasize physical processes, nutrient cycling, and speciation of metals (Evans *et al.*, 1997).

Until recently, there has been little standardization for methods of sediment collection, handling, transport, storage, and sample preparation. That has been remedied by a guidance manual (EC, 1994c) which outlines procedures preparatory to biological testing or chemical characterization, and differentiates between procedures for research and regulation.

Commonly used tests. Environment Canada has several standard methods for testing the toxicity of sediment. For marine and estuarine habitats, a 10-d test measures the mortality of amphipods (EC, 1992e) and a 14-d test assesses mortality and growth in the polychaete worm *Polydora cornuta* (EC, in prep., b). A 14-d exposure is used to evaluate mortality and growth of the freshwater amphipod *Hyalella azteca*, and this test can also be used for estuarine conditions (EC, 1997b). Freshwater sediments can also be evaluated with a 10-d test of the mortality and growth of midge larvae (EC, 1997a). All these tests are similar to methods used elsewhere. There are continuing improvements in details of apparatus or technique, for example, exchange of overlying water in the test chambers can be accomplished with reduced effort by means of devices that are relatively simple (Leppanen and Maier, 1998) or complex (Wall *et al.*, 1998).

Newer sediment tests emphasize sublethal effects including reproduction, in life-cycle or at least chronic exposures. Such tests can be done in 28-d tests with either marine (Emery *et al.*, 1997) or freshwater amphipods (Ingersoll *et al.*, 1998). A life-cycle exposure measures reproductive success of a freshwater midge in about 65 days (Sibley *et al.*, 1996; Benoit *et al.*, 1997). Lee and Tay (1998) show how bacterial activity can be used to assess sediments for environmental impact assessments; they, and Ross (1998) point out how sediment testing can be (and should be) broadened to include this important group of organisms. Small polychaete worms that can be cultured in the laboratory are being developed as test organisms suitable for Canada's Atlantic and Pacific coasts (Pocklington and Doe, 1998).

4.11.1 Tests with Whole Sediment

Key Guidance

- *Tests with whole sediment are commonly used to establish the toxic status of bottom deposits.*
- *Multi-concentration tests with whole sediment can be done by adding test sediment to the control sediment; procedures are not yet completely standardized. Chemicals can also be used to spike the control sediment.*
- *Sediments can have natural toxicity. Ammonia is the most common cause, while hydrogen sulphide and lack of oxygen can also be toxic.*
- *A **control sediment** is used for quality control, is known to be nontoxic to the test organism, and can be natural or artificial.*
- ***Reference sediments** should be as similar as reasonably possible to the physicochemical characteristics of test sediments, but with less anthropogenic contamination.*
- *Particle size is the most important variable in selecting reference sediments. Content of organic matter is also important, while qualitative characteristics of clays and organic matter, and salinity, are secondary considerations.*
- *Ideally, reference sediments should be collected near the site of the test sediments, outside the zone of contamination, to improve likelihood of comparable characteristics. It can be a problem to match sample characteristics during a field survey, before any physical and chemical profiles of test sediments have been done. One tactic is to collect a series of reference sediments bracketing the characteristics of test samples, another is to formulate a matching control sediment which can be used as a reference sediment.*
- *Comparison of biological responses to control and reference sediments might reveal toxicity from natural sources or from unsuspected contaminants.*

Whole-sediment toxicity tests are commonly used for assessing the quality of sediment samples from the field. The most direct tests expose organisms to samples of sediment that have not been fractionated (i.e., no division of the sample by extracting porewater, nor by segregating particle sizes, etc.). The tests use an endpoint of percent effect such as percent mortality). Testing commonly involves full-strength whole sediments, perhaps to map toxic and non-toxic zones of sediment in a harbour. Alternatively, graded concentrations of sediment(s) might be tested (see "Spiking"). Complete interpretation of either kind of test would require both *control sediment* (known to be clean) and *reference sediment* (natural, thought to be relatively clean, see following text).

Natural toxicity. Toxicity of a sediment sample is not necessarily an indication of human contamination because there can be high levels of natural toxic substances. Therefore, reference sediments are critical for interpreting toxicity tests, and their results should be carefully appraised. A chemical profile of the test sediment might be sufficient to rule out most common natural toxicants, but parallel toxicity tests on reference sediment can help to estimate the relative contributions of natural and/or anthropogenic toxicants.

Ammonia is the most common natural toxicant, and is found in organically enriched sediments with active decomposition (Ankley *et al.*, 1990; Burton, 1991; OECD, 1992c). Sarda and Burton (1995) found that ammonia in sediment increased during the warmer months, and showed appreciable spatial variation at a given site. Even measuring the ammonia can be challenging. The same authors found higher concentrations in samples collected by Ekman grab and coring device, compared to porewater samples collected by suction. Changes were also observed during storage of porewater for more than one week.

Natural ammonia toxicity could be present in the samples of sediment being tested for anthropogenic contamination, confounding the results. Ammonia was the single cause of toxicity, or most of it, in a Belgian river (Van Sprang and Janssen, 1997), and had a probability of up to 18% of being the only cause of toxicity in a series of 322 marine sediments in the eastern USA (Moore *et al.*, 1997). Ammonia toxicity in sediment tests is reviewed by Tay *et al.* (1998), who point out the profound effects such background toxicity could have for regulatory decisions on marine disposal of dredged waste. They conclude that the investigator must make a decision on whether to include ammonia toxicity in the test result, or to attempt to eliminate the ammonia problem by aeration or other means. In either case, the investigator should at least understand the contribution of ammonia and other nature toxic agents.

Hydrogen sulphide is also a common source of extraneous toxicity in sediments (Jacobs *et al.*, 1992; Knezovich *et al.*, 1996; see also Section 4.11.2). Anoxia from decay of an enriched sediment can be deleterious to organisms, and such anoxia is, indeed, almost standard below the surface layer of a sediment (Krantzberg and Boyd, 1992). Cyanide can be a problem. Other toxic byproducts of anaerobic decay are known, and even undecomposed plant defence compounds might contribute toxicity.

Spiking. A multi-concentration series must be created if it is desired to estimate the LC50 or EC50. This can be done by spiking aliquots of control sediment with appropriate amounts of a chemical of interest, or by adding suitable amounts of a contaminated test sediment. Chemicals may be added in a carrier solvent (organic substances) or in aqueous solution (metals), followed by various mixing techniques (EC, 1995a). Detailed guidance on spiking sediments is given in the Environment Canada guidance document (EC, 1995a). Laboratory-spiked sediments are considered representative of sediment that has recently been contaminated. Sediments with a long history of contamination might have toxicant molecules more strongly bound to surfaces of sediment particles, thereby reducing their bioavailability (Burton, 1992).

Spiking with a toxic substance might be used as a tool in attempts to identify causes of observed toxicity. For example, testing of sediments near New York suggested that dioxin contamination of sediments might be responsible for toxicity. Spiking a reference sediment with dioxins, however, showed no toxicity at any of the concentrations, indicating that other causes should be sought (Barber *et al.*, 1998).

Techniques for diluting with contaminated test sediments are not yet standardized, and many difficulties are found in achieving homogenous and representative mixtures. For example, the physical structure of the sediment can be disrupted by mixing it with the control sediment, or the control might be different from the test sediment in such critical variables as grain size and chemical binding capacity. Hence, mixing of the two sediments can change the physicochemical properties and confound the findings of the toxicity test [40].

---

[40] Adding clean sediment to the test sediment is not the same as simple dilutions in water-based tests. The clean sediment provides fresh reactive surfaces to adsorb toxicants from the test sediment, reducing bioavailability to a greater degree than would be estimated from the simple proportions of the two kinds of sediment (Burton, 1992). Clean sand has been used in tests because it has low adsorptive capacity and is less likely to reduce bioavailability (e.g., Tay *et al.*, 1992). Sand has, however, resulted in U-shaped response curves in multi-concentration toxicity tests, with lowest apparent toxicity at

Because of non-standardization and the difficulties previously mentioned, there remains some freedom in choosing a method for diluting a test sediment into another "clean" sediment. However, the best choice for the "clean" sediment would be a formulated or artificial control sediment, made up to mimic the test sediment in organic content, range of particle size, and proportions of sizes. Choice of a formulated control sediment is supported by information in Environment Canada methods documents on sediment tests with freshwater amphipods, midges, and marine polychaetes (EC, 1997a; b; in prep., b), and in the spiking document (EC, 1995a). Uncontaminated sediments from the same area as the test sample (i.e., reference sediments) would also be useful, but only if they were similar in grain size and other characteristics that influenced chemical activity (Nelson *et al.*, 1993).

Control sediments. These are natural sediments or artificially constituted ones ("formulated"), of defined composition and consistent quality, known to be without direct chemical toxicity for the test organism. Frequently, control sediments are those used in culturing test organisms (Burton, 1992). Control sediments are an integral element of quality control for toxicity tests with sediment (EC, 1995a). Use of several control sediments with different physicochemical properties (e.g., particle size or organic content) allows a comparison of biological effects in the test samples (caused by toxicants and/or the physicochemical properties) with any effects in the controls caused by properties unrelated to contaminants. Comparison of biological responses in control and reference sediments (see following text) might reveal whether the reference had chemical toxicity from undetected contaminants, or from chemicals naturally present, or might distinguish effects due to the differing physicochemical characteristics.

Reference sediments. These should be collected near the site of the test sediments, but outside the zone of contamination. The reference sediment is presumed to be clean. Reference sediments are chosen so that physicochemical characteristics are as similar to test sediments as can be practically achieved within the general locale of the area under investigation (Burton, 1992). The chief characteristics to match would be similarity in particle-size distribution, organic content, the kinds of clays and organic matter, and the salinity of porewater. Investigators report difficulty in attempting to match TOC levels in reference and test sediments, since TOC is often higher in polluted areas.

In sediments, the particle size (grain size) influences the mobility and behaviour of toxicants. Organic substances and metals bind more readily to fine silts and clays than to coarser sand and gravel, because the fine particles have more surface area per volume, and because they have greater surface charges. Fine sediments include clays, metal oxides, sulphides, and organic matter aggregates (Lamberson *et al.*, 1992). Increasing proportions of fine sediment can cause major increases in measured toxicity in the Microtox test for sediment (Tay *et al.*, 1998). Grain size can also be a major habitat factor for benthic invertebrates used as test organisms. Survival can be low in sediment that has unsuitable grain size, even in the absence of contamination (DeWitt *et al.*, 1988; Becker *et al.*, 1990). Dense sediments can impede burrowing, tube-building, or feeding, while coarse sediments might not provide support, food or allow tube-building (Nelson *et al.*, 1993). These effects might reduce toxicity, if the organism did not burrow, or increase it by imposing a secondary stress (Lamberson *et al.*, 1992). No simple generalization can be made for invertebrates, and Tay *et al.* (1998) describe the findings that in sediment tests with different amphipods, some species showed a major effect of particle size, while others showed little or no effect.

Porosity is the extent of the network of openings in sediment, which allows the passage of water into and through the sediment. Porosity is determined by the particle size distribution, but also decreases with depth in the sediment because of mass compaction. Porosity is one factor that governs the water-holding capacity of the sediment, and it also mediates the quality of porewater. More porous sediments contain greater volumes of porewater, but will tend to be less toxic on a volume basis because of dilution. Porous sediments allow more rapid downward migration of oxygen from overlying water, and more rapid upward migration of toxicants from anoxic layers (chiefly ammonia). Porosity also influences habitat preferences of burrowing organisms.

Standardization of the organic carbon and particle size in a reference sediment has a broad influence in stabilizing other attributes. Content of organic matter (measured as

---

intermediate concentrations (DeWitt *et al.*, 1988; Nelson *et al.*, 1993).

total organic carbon, TOC) largely determines the complexing capacity of the sediment for metals, organic substances and natural organic toxicants. Accordingly, normalization for TOC can be used to correct for differences in complexing action (Section 4.11.3). Similarly, many physical and chemical attributes tend to be correlated with particle size, and become more comparable if grain sizes are standardized.

If the reference sediment is taken from a site that is nearby and similar to that of the test sediments, it is more likely to be similar to the test sediments. For example, if nearshore river sediments below a city were being tested, appropriate reference sediments would be from the same shore upstream of the city, rather than at mid-river adjacent to it, where environmental influences would likely differ. For a marine disposal area, reference sediments should come from a clean area in the same bay, sound, harbour, or cove, or if necessary from a similar embayment close by. Samples should come from approximately the same depth, and from a location with the same relationship to major influences on sediment quality (e.g., currents, river inflows). Choice of location in estuaries should avoid confounding differences in salinity of porewater. Some investigators collect the test sediment(s) first, then do exploratory sampling and collect three or four potential reference sediments. These are assessed for toxicity and chemical characteristics in the laboratory, and the most appropriate one is selected for the testing program.

Practical difficulties in collecting reference sediments, and how to deal with them, are covered in Section 5.3.2.

If a suitable reference sediment cannot be obtained, a control sediment might be formulated with TOC and grain size matching the test sediment. However, this option requires physicochemical appraisal of the test sediments before the formulation can be designed. Hamr *et al.* (1994) have studied formulation of artificial reference sediments in relation to the Environment Canada toxicity tests. There is promise that increasing knowledge may allow accurate prediction, from the physicochemical characteristics of a sediment, of its suitability for various qualities of biological community. Such knowledge of "reference conditions" (Day *et al.*, 1996) should assist in defining suitable reference sediments in toxicity testing (see Section 5.5.4).

4.11.2 Tests with Porewater or Extract

Key Guidance

- *Toxicity of porewater from a sediment is a useful measurement because it is often found to be equivalent to the toxicity of all the water-soluble contaminants of the whole sediment. That empirical observation is supported by expectations from research on partitioning of nonpolar organic contaminants between the solid phase and porewater.*

- *Toxicity tests on the whole sediment tend to be favoured by experts, as being more representative of the true situation. This is partly because a porewater test oxygenates the porewater which might normally be anoxic, a process that also lowers or eliminates concentrations of toxic sulphides, and changes metals from non-toxic to toxic forms.*

- *Tests using solvent extracts of sediments tend to overestimate toxicity of the sediment, and are most useful for exploratory surveys and in identifying the nature of toxic components.*

Two useful types of liquid samples can be generated from sediments in the laboratory.

1. **Porewater** samples collected by physical treatment of sediments, usually by centrifuging and filtering.

2. Non-aqueous **extracts** are prepared by adding a nonpolar solvent and shaking to move bound contaminants into the solvent. The solvent can be used as a carrier in aquatic toxicity tests or removed by evaporation (Schiewe *et al.*, 1985)[41].

---

[41] As an example, nonpolar solvents have frequently been used for extractions from sediment in preparation for the Microtox™ liquid-phase bacterial luminescence test (Ankley *et al.*, 1989; Tay *et al.*, 1992; Jacobs *et al.*, 1992; Ho and Quinn, 1993).

Different sediments that have the same total concentrations of organic compounds or metals can have very different toxicities because they bind chemicals to different degrees, influencing bioavailability (Di Toro *et al.*, 1990; Giesy *et al.*, 1990). For benthic organisms, the principal route of exposure for some chemicals is through sediment porewater, rather than from the solid phase (Di Toro *et al.*, 1991), although there are important exceptions [42]. The concentration of chemical in the porewater is chiefly determined by its affinity for certain solid components, especially organic carbon. Thus, an equilibrium of partitioning is established between the fraction of the chemical sorbed to sediment particles and that dissolved and available in the porewater.

Equilibrium partitioning is achieved when the chemical *activity* of a toxicant is equal in the aqueous and solid phases, although mass concentrations might be very different. Di Toro *et al.* (1991) provide a comprehensive review of equilibrium partitioning. The most easily followed description of numerical procedures is given in OECD (1995a; Section 9).

Thus, porewater concentrations indicate the partitioning behaviour of chemicals within sediments (Ankley *et al.*, 1991a). It follows that toxicity of porewater should often represent toxicity of the whole sediment. If that were always true, there would be a gain in convenience of testing, since multi-concentration tests could easily be performed, and if desired, well known aquatic tests could be used (e.g., daphnid or fathead minnow tests, which are well-standardized although they do not use sediment-dwelling organisms). Carr (1998) reviews the advantages of tests with porewater, and also the suitable organisms and practical matters in carrying out the tests.

There are, however, other factors which can make porewater tests less than adequate as an estimate of whole-sediment toxicity, and both aspects of the question are described in the following text. At present, investigators assessing the toxicity of sediments should likely use a combination of methods for both porewater and whole sediment.

There is some agreement that testing with porewater can be useful and representative of whole-sediment toxicity. Porewater tests can slightly overestimate sediment toxicity (Sasson-Brickson and Burton, 1991; Ankley *et al.*, 1991b), although there are exceptions (Giesy *et al.*, 1990). Detailed guidance on collection of porewater, *in situ* and in the laboratory, is provided in Environment Canada (1994c) and in Environment Canada test methods for sediments (EC, 1992e; EC, 1997a, b; in prep., b).

Despite the apparent usefulness of porewater, there does not seem to be consensus among experts, that its use can replace testing with whole sediments. One reason for this concerns the normally anoxic conditions in subsurface sediments, and the presence of sulphides. As the result of bacterial reduction of sulphate, the interstitial water of sediments can contain hydrogen sulphide (H_2S) and sulphide ion (HS); the former dominates below pH 7 and the latter above pH 7. Both are easily lost within a day or two by aeration and oxidation during the extraction and testing of porewater. Their toxicity is not, therefore, reflected in porewater tests. In the undisturbed sediments, the sulphide compounds tend to detoxify metals by forming metal sulphides and polysulphide complexes; nontoxic metals can include copper, cadmium, mercury, lead, and zinc. Upon oxidation, the metals change their speciation and can show toxicity in the porewater test (Wang, 1998; see also Section 4.11.4).

Although *elutriates* are used in testing soils, they are not appropriate for testing sediments, nor does Environment Canada suggest that they be used.

Solvent extracts tend to be more toxic than porewater (e.g., Tay *et al.*, 1992) but they are more difficult to relate to potential effects in the environment. Solvents can mobilize particle-bound toxicants that would have low biological activity in nature, and toxicity of the solvent itself can distort test results (see Sections 4.12.3, 4.13.1).

---

[42] Hydrophobic organic compounds and some metals are more strongly bound to particles and exposure to the particles is more important for the organism. For example, benthic organisms which feed by passing sediment through the gut, accumulate benzo(a)pyrene and hexachlorobiphenyl mostly from the ingested particles (Kukkonen and Landrum, 1995). Other substances can show greater bioavailability from whole sediment than from pore water or elutriates (Harkey *et al.*, 1994). For some deposit-feeding animals, uptake from the gut may be much more important than uptake through the integument from porewater; 20 to 30 times as much fluoranthene was taken up through the gut by the marine polychaete *Capitella* (Forbes *et al.*, 1998).

Apparent toxicity of the same sediment varies according to the solvent used (Ho and Quinn, 1993), and natural toxicants such as sulphur can be extracted along with anthropogenic contaminants (Jacobs *et al.*, 1992). Solvent extracts are most useful for preliminary surveys to establish the location of potentially toxic sediments, and when attempting to identify the toxicants present in a sediment (Section 4.7).

4.11.3 Normalization for Total Organic Carbon

Key Guidance

- *The concentration of an organic chemical that is dissolved in the porewater and biologically available can be estimated by $K_{oc} \times f_{oc}$, where K_{oc} is the partition coefficient for organic carbon, and f_{oc} is the proportion of organic carbon in the sediment.*
- *This equation can be used to normalize toxicant concentrations in sediments that differ in their organic carbon, for comparison of potential effects. The relationships are not yet well documented, however, and different kinds of organic carbon with different sorption capacities may limit the improvements in normalizing toxicity.*

Partitioning between solid and aqueous phases evidently determines bioavailability of many organic compounds and metals in sediments. Total organic carbon (TOC) is the chief property of sediments that influences the partitioning of nonpolar (neutral) organic compounds (Di Toro *et al.*, 1991). Humic and fulvic acids make up most of the organic matter in sediments, and TOC is an indicator of their concentration. These polycyclic organic acids have a high affinity for hydrophobic organic compounds, with which they form stable, chemically inactive complexes, and which in many cases are also biologically inactive.

Generally, as the TOC content of a sediment increases, greater proportions of any given chemical in the sediment will be bound to organic matter and rendered biologically unavailable. Thus, different sediments containing the same mass of contaminant can show very different toxicity, in relationship to the TOC concentration (Nebeker *et al.*, 1989). Accordingly, TOC can be used to normalize contaminant concentrations according to their potential for biological activity (Di Toro *et al.*, 1991).

The affinity of an organic compound for the carbon compounds in sediment is determined by its hydrophobicity, measured by its *organic carbon partition coefficient* (K_{oc}). The value of K_{oc} is predictable from regression equations using octanol–water partition coefficient (K_{ow}) (Section 4.6), and for practical purposes the two are equivalent for sediments (Di Toro *et al.*, 1991). The fraction of dissolved, biologically available chemical as a proportion of the total chemical mass in the sediment is predicted by $K_{oc} \times f_{oc}$, where f_{oc} is the proportion of organic carbon in the sediment. (The porewater concentration includes some chemical bound to colloidal humic acids, but this is also allowed for in the partition coefficient).

For toxic substances in sediment, the estimated concentration available in porewater should be a reasonable basis for predicting biological effects. For a number of species and neutral organic compounds, acute lethality in toxicity tests with sediments occurs at concentrations very close to those for standard aquatic tests, after normalizing for sediment TOC (Di Toro *et al.*, 1991). Equilibrium partitioning based on TOC has been proposed as a method of establishing sediment quality criteria from water quality criteria (Di Toro *et al.*, 1991).

This rationale seems reasonable, but is probably overly simplistic. All organic carbon is not equal in the sorption of organic toxicants. Carbon normalization does not necessarily account for all the differences among sediments, and might not greatly improve prediction of toxicity, beyond that obtained by gross chemical measurements (Landrum and Faust, 1995; Kukkonen and Landrum, 1996). Further clarification can be expected through current research.

Metals can also partition to organic carbon in sediments. Mahony *et al.* (1996) investigated this and show that the relationships for copper and lead conform to the theoretical model. The measured activity of cadmium, however, did not conform to theoretical binding, and no explanation was evident.

4.11.4 Normalization for Acid-volatile Sulphide

Key Guidance

- *Certain metals (Cd, Cu, Ni, Pb, and Zn), can be bound by sulphur in the sediment and become unavailable. This partitioning between the solid phase and sediment porewater can be assessed in most sediments by measuring acid-volatile sulphide. Toxic effects of the metals would not be expected if the molar ratio of metal to sulphur in sediment were less than 1.0.*

The partitioning of some metals between sediment solids and porewater can be estimated by measuring the content of acid-volatile sulphide (AVS) of the solid phase (Di Toro *et al.*, 1990; Carlson *et al.*, 1991). AVS is the fraction of solid-phase sulphur in sediment, mostly iron monosulphide (FeS), that is extractable in cold acid. Other cationic metals can displace iron from FeS to form insoluble monosulphides, which remain in the solid phase and are not biologically available. A complex such as zinc sulphide appears to be more stable than the iron-sulphide complex that it replaces, as would be expected in theory (Liber *et al.*, 1996). The ratio of metal to sulphur in these sulphides is 1:1, so AVS will not scavenge metal ions from the aqueous phase, once the molar ratio of metal to S has risen to unity (1.0). Metal measurements are conventionally described as *simultaneously extracted metals* (SEM), and molar ratios are signified by SEM:AVS. If the ratio were ≤1.0, metal would be expected to be unavailable to cause toxic effects. In some trials, this system predicted lack of toxicity with 99% accuracy, and predicted toxic sediments with 79% accuracy, when combined with appraisals of total metals in the sediments by the toxic units method (Hansen *et al.*, 1996). The predictive method is also valid for sublethal effects in longer tests (Sibley *et al.*, 1996). Bioaccumulation of divalent metals from sediment is usually predictable from the AVS/SEM ratio, but there are occasional exceptions. Research has improved the usefulness of AVS for predicting bioaccumulation and toxicity of metals (Ankley, 1996; Berry *et al.*, 1996); investigators concerned with metals in sediments would be well advised to follow current publications.

In theory, any cationic metal whose monosulphide is less soluble than FeS should displace Fe in sediment AVS, to produce insoluble metal monosulphides. In practice, there are several limitations to the relationship.

- At present, only cadmium, copper, nickel, lead, and zinc are included as metals which can be removed from bioavailability by AVS. (It is possible that arsenic, chromium, silver, and mercury might show a similar detoxification by AVS (Di Toro *et al.*, 1990).
- There is no guarantee that metals *will* be bioavailable if SEM:AVS >1.0; other factors can come into play [43].
- Very low concentrations of SEM and/or AVS should not be considered predictive. Clearly, AVS should be above operational detection limits to be meaningful. Di Toro *et al.* (1990) indicate that the relationship is not meaningful for oxidized sediments in which the AVS is <1 μmol/g. Measurable but low concentrations of AVS signify that the absolute values of SEM and AVS should be considered as well as the ratio. Relatively low concentration of AVS with somewhat more SEM should not be given the credence of high concentrations at the same ratio.

Empirical observations have almost always confirmed that SEM/AVS ratios below 1.0 are associated with no toxicity from the five metals cadmium, copper, nickel, lead, and zinc. (Of course, there might be other substances present which would make a given sediment toxic.) It appears that more insight into binding capacity, and the effects on organisms, can be obtained by using a subtraction (SEM - AVS) rather than a ratio (Hansen *et al.*, 1996). In that case, values below zero would indicate no toxicity.

In freshwater lakes the sediment AVS levels can vary seasonally, with highest concentrations in deep parts of the lake during the anoxia of summer stratification (Howard and Evans, 1993). That variation shows the importance of considering seasonal effects when making evaluations in a field program. It also shows how important it is to retain the field characteristics of any sample transported to the laboratory for testing. AVS is labile in oxic sediments, and care must be taken to maintain anaerobic conditions when sampling and transporting sediments and during the AVS extraction (EC, 1994c). (The issue of suitable oxygen for the organisms when testing porewater is considered in Section 4.12.2).

---

[43] At low AVS concentrations, other constituents, in particular humic and fulvic acids (TOC) as well as iron and manganese oxides, can become important influences on metal partitioning (Di Toro *et al.*, 1991; Fu *et al.*, 1992).

4.12 *Toxicity of Soils*

> ***Key Guidance***
>
> - *Toxicity testing of soils is relatively new. Terrestrial ecosystems have unique features, and interpretation cannot simply be extrapolated from aquatic systems.*
>
> - *Physical structure and chemical nature of the soil influence test results strongly. Proper controls and/or reference tests are vital for correct interpretations.*
>
> - *Soils contain all three primary media, i.e., solids, water, and air. Various contaminants can reside in any of these media, and can use one or more of them as a pathway for reaching and affecting living organisms.*

Toxic substances enter soils through disposal of wastes, atmospheric fallout, and direct applications, either accidental (spills) or deliberate (pesticide spraying). Land used for industrial purposes has commonly become contaminated with metals, PAHs, volatile organics, and other organic compounds (Dasappa and Loehr, 1991). Widespread, low-level soil contamination also arises from such ubiquitous sources as oil leaks and tire wear from cars, and storage of erodible, leachable, or corrodible substances in direct contact with soil. Metals and hydrophobic compounds tend to bind to soil particles, and once bound, the organic compounds often resist biological decay. Even volatile organics can remain in soils for many years if they migrate below the zone of biological activity. Consequently, soils tend to be a net sink for contaminants, or if disturbed they can be a source of contamination via leaching or runoff.

Terrestrial ecosystems can be seriously damaged by contaminated soils, since soils are their supporting foundations. The entire food web ultimately depends on plant growth, and the plants depend on soil for anchorage, water, and nutrients. Toxicants are also a direct threat to soil-dwelling invertebrates and microbes, and can inhibit essential processes such as decomposition and nutrient cycling.

Contamination from a spill can migrate toward groundwater, and the speed and direction depends on the affinities of the chemical for water or soil particles, and the volume and prevailing flow of groundwater. Substances tend to be attenuated as they move through the soil. Consequently, a massive acute exposure at the soil surface can become a chronic sublethal concentration in groundwater and perhaps in down-gradient waterbodies.

Testing the toxicity of soils is a relatively new field, with methods and interpretation less developed than for testing effluent, water, and sediments. Tests for soils and aquatic sediments (Section 4.11) share many features, but terrestrial ecosystems are distinct and interpretive approaches cannot be extrapolated from aquatic ecosystems. Environment Canada provides a comprehensive compilation and comparison of test methods for contaminated sites (EC, 1996a). Comprehensive information from Europe is given in a handbook of sublethal toxicity tests for soil, using invertebrates (Løkke and van Gestel, 1998). The first part of that handbook gives background on approaches for testing soil, while the second part describes and discusses ten test methods and organisms. Proceedings from a workshop on testing with earthworms are presented by Sheppard *et al.* (1998).

The physical structure and chemistry of the soil have strong influences on the results of toxicity tests. Organisms in soil can be exposed to toxicants from three phases: the vapour in pore spaces; the porewater; and the solid particles themselves. Soil structure and the nature of the phases vary with locality, so knowledge of the pathways of a toxicant through these phases is obviously important for interpreting and forecasting impact. Such knowledge is being gathered for metals, for example, by using aqueous toxicity data and the appropriate solid-liquid partitioning coefficients. That implicitly assumes that (a) concentration of a metal in porewater is the important factor, toxicologically; and (b) there is chemical equilibrium between the concentration in porewater and the solid phase (Janssen *et al.*, 1997a, b). The assumptions appear reasonable, since the bioconcentration of metals into soil organisms appears to be governed by the same characteristics as the partitioning coefficients between soil and porewater (Janssen *et al.*, 1997b). For copper and lead, Sauvé *et al.* (1998) review the estimation of free metal activities in porewater by means of regressions based on chemical speciation in contaminated soils. In turn, the free metal activity can be related to toxicity, opening an approach for derivation of criteria for soil quality.

Interpretations of toxicity results for soils are greatly aided by proper controls and reference tests. Extrapolation of any results to the environment requires caution, and guidance is given by Sloof *et al.* (1986).

A final complexity is that biodegradation of the contaminating substance is particularly important in testing soils. Approaches for evaluating degradation in soil, and including that in environmental evaluations, have been developed for categories of synthetic polymers, and described in several chapters of a book on polymers (Hamilton and Sutcliffe, 1997).

4.12.1 Tests with Whole Soil

Key Guidance

- ***Control soils** are natural or artificial soils that are known to be nontoxic to the test organism, and their use is an element of quality control. A **reference soil** is a natural soil selected to have physicochemical characteristics similar to the test soil, but has little anthropogenic contamination. Comparisons with control and reference soils can reveal stress from natural sources such as unfavourable soil structure, or undetected contaminants.*
- *A series of concentrations of a toxic test soil can be created by diluting it with a control soil, often an artificial one.*
- *The potency of a known toxic substance in the soil of interest can be tested by spiking the substance into subsamples of the soil, and testing the toxicity.*

The simplest and most direct toxicity tests with soil expose organisms to largely undisturbed samples of whole soil. Whole-soil tests are commonly used to screen field-collected samples, with an objective of overall assessment of soil toxicity. Such a test might measure growth and survival of earthworms, or seed germination and emergence. Toxicity is commonly expressed as the percentage of earthworms surviving, their mean percent weight loss after a fixed period of time, or the percentage of seeds which develop shoots.

Whole-soil tests are best done by comparing the results with those of both control and reference soils. A *reference soil* has physical and chemical characteristics similar to the test soil, but without appreciable anthropogenic contamination, and will usually be collected near the contaminated area (Section 4.12.2).

Control soils are natural or artificially constituted soils of defined composition and consistent quality, and are known to be without direct chemical toxicity for the test organism. Frequently, they are artificial and their composition is specified in the test method. Control tests show whether biological effects arise from background sources such as unhealthy test organisms or poor quality dilution water, which might be contributing to the *apparent* toxicity of test samples. By indicating such "baseline" effects, control soils serve a purpose for quality control. Comparison of findings for control and reference soils can reveal whether the reference has chemical toxicity from undetected contaminants, or from chemicals naturally present.

Multi-concentration tests. A series of concentrations of the whole soil is required to estimate a quantitative endpoint (LC50, EC50 etc.) instead of percent response in full-strength soil. That might be desirable, for example, to map the various degrees of contamination at a hazardous waste site. There can be procedural problems in diluting test soils. The physical structure of the soil can be disrupted by mixing it with the diluent, or the diluent might be different in grain size, chemical binding capacity, or other important characteristic. Hence, diluting whole soils can change the physicochemical properties of the soil matrix and confound results of any subsequent toxicity tests. Two remedies can be used, (a) artificial soils, and (b) liquid extracts (Section 4.12.3).

An artificial soil (i.e., a formulated control soil) is often used for multi-concentration tests. It acts as the test medium, and allows effects of toxicants to be separated from effects of soil type. The standard diluent facilitates comparability among different toxicity tests. If physical characteristics of the test soil change upon dilution, the nature of the change should be documented.

Most artificial soils attempt to mimic a typical sandy loam, and consist of a mixture of sand, clay, and peat, to create physicochemical characteristics conducive to the health of plants and invertebrates. Simpler media have also been used. For example, pure sand can suffice for testing seed germination because it need only provide anchorage for the seeds, and has relatively few adsorptive surfaces to bind chemicals (Baud-Grasset *et al.*, 1993). Using a standard soil makes it more difficult to extrapolate results to a different soil type in the environment. In Canada, a standard sandy loam might not be at all representative of the local types that are of concern, such as forests or tundra.

A reference soil could be advantageous for dilution in Canadian situations, because of the likelihood of a close match to the test sample. A very good match might be critically important for predictions at a hazardous waste site.

If the toxic chemical or substance is known, its action in a particular soil can be assessed by spiking various concentrations into a series of subsamples of the whole soil. Multiple concentrations would thus be created for a toxicity test. Thorough mixing would allow the substance to bind or otherwise react with the soil particles. Such tests with spiked chemicals would represent toxicity for recent entry into the soil. In the longer term, surface-bound molecules might be more strongly bound, less bioavailable, and accordingly less toxic. In such tests, toxic organic substances would usually be added in a carrier solvent, and metals in an aqueous solution.

4.12.2 Selecting Reference Soils

Key Guidance

- *Reference soils should be collected near the location of test samples, but outside the zone of contamination. They should be as similar to the test soils as possible, especially in texture (grain size), porosity (cavities), organic matter, and cation exchange capacity.*

- *Reference soils should be from the same Order (e.g., chernozem) and Great Group (e.g., dark brown chernozem) as the test soils. They must come from the same horizon (level) of origin, and must have suffered the same level of disturbance.*

Reference soils should be as similar to the test soils as possible, in physical and chemical characteristics. The most important similarities are in texture (grain size), porosity (cavities), organic matter content, cation exchange capacity, and horizon of origin. Soil science texts give detailed information on the properties of soil (Foth and Turk, 1972; Brady, 1974).

The texture of a soil governs its other characteristics, the health of endemic organisms, and the activity of toxicants. Coarse, large-grained soils drain rapidly, have low nutrient content, and low capacity to bind polar compounds and ions. Conversely, soils high in fine silts and clays have less internal space, more capacity to retain moisture, and a higher binding capacity for charged substances including nutrients and organic compounds. Soils high in clay are slow to drain and can be too dense for burrowing organisms.

Soil porosity is closely related to texture. Porosity refers to the size and inter-connection of open spaces among solid particles, and is determined by texture, compaction, and organic forces like burrowing animals and plant roots. The spaces fill with water or air depending on moisture content of the soil. At field capacity (i.e., the moisture content of soil when all gravity drainage is finished), coarse soils are lower in water content than soils of fine texture. Porous soils drain quickly and allow more rapid migration of contaminants. Texture/porosity also has a major influence on the burrowing ability of organisms such as earthworms, or on the penetrating ability of plant roots.

Clays bind some toxicants. Particles that have a net negative charge can bind cations, according to the density of exchange sites measured by the cation exchange capacity (CEC). This is important agriculturally, but also influences the binding of metals and polar organic compounds. Clay type and content affect CEC. Multi-layered smectite clays have a higher and stronger binding capacity than simpler kaolinite clays, making polar substances less available.

Organic matter has a greater complexing capacity than the clays, per unit weight, and is a major factor in binding of both polar and nonpolar substances by soil. Hydrophobic organic compounds bind strongly with particulate organic matter and show extremely low bioavailability. Organic content is also an important habitat variable for soil organisms, and influences other features like pH and microbial activity. Organic matter of agricultural soils is <2% by weight (except for muck soils), but undisturbed range and forest soils can have up to 20–30%.

Distinct horizontal layers *(horizons)* develop naturally in soils because of downward elution and deposition on top. Structure and chemical behaviour can be very different between horizons, so it is important that reference and test samples come from the same horizon. Most toxicity tests with soil are concerned with the top (A) horizon, but B or C horizons might be involved in deeper contamination.

Soil classification in Canada involves *Orders* (e.g., chernozem) and *Great Groups* (e.g., dark brown chernozem) (Agriculture Canada, 1987). Reference soils should be from the same Order as the test soil, as a minimum requirement. If taken as close together as possible, they will likely be of the same Great Group. The reference soil must also have suffered the same degree of disturbance (e.g., agricultural soils should not be paired with undisturbed forest or rangeland soils). If contamination is widespread and a reference soil must be collected off-site, it can still be comparable if it matches the test sample in its Great Group, horizon, and degree of disturbance.

4.12.3 Tests on Soil Extracts

Key Guidance

- *Leachates obtain the most water-soluble substances from the soil, and can miss some contaminants that are slowly mobile.*
- *Elutriates are considered reasonably representative of the true plume of substances that might enter groundwater.*
- *Solvent extracts have been used to estimate the total quantity of toxic contaminants in a soil sample, but the validity of that assumption remains uncertain.*
- *Tests using whole soil with terrestrial plants and invertebrates continue to be the best benchmark of environmental hazard.*

Liquid extracts of soils can be used in toxicity tests with some apparent procedural advantages. First, they allow non-confounded multi-concentration tests on the extracts. Second, better-known standard aquatic tests can be used. This is a questionable advantage because of the difficulties in interpreting the ecological significance to soil organisms, of results obtained from strictly aquatic organisms. The rationale for using aquatic tests would be more convincing if it were desired to assess the hazard from groundwater that might re-surface into aquatic ecosystems (Section 5.5.2). Toxicity of liquid soil extracts is commonly evaluated with bacterial luminescence or other bacterial tests, which provide useful rapid diagnosis. Tests based on seedling root-elongation would also be suitable.

Three main types of liquid samples extracted from soils are used for toxicity tests.

1. *Leachates* formed by water passing through a column of soil, by gravity with no agitation.
2. *Elutriates* produced by mixing defined amounts of soil and water (usually 4 mL water per gram of soil) with agitation to obtain soluble and weakly bound constituents. Solids are then removed by centrifugation and filtration or by decanting (Donnelly *et al.*, 1991).
3. *Non-aqueous extracts* using a nonpolar solvent such as methanol, methylene chloride, or dimethyl sulphoxide, and mixing to obtain most of the bound contaminants. The solvent is used either as a carrier in toxicity tests or else it is evaporated. Extracts are commonly tested by bacterial luminescence or similar bacterial tests (Donnelly *et al.*, 1991).

Detailed guidance on sampling soil and preparing extracts is provided by Environment Canada (EC, 1996a). Advice in EC (1996a) is useful because the literature on the various kinds of liquid extracts from soils tends to be inconsistent and confusing.

Leachates collect more water-soluble substances than other methods of extraction, but might not obtain some moderately hydrophobic substances that could slowly migrate down-gradient to nearby waterbodies. Dasappa and Loehr (1991) consider that elutriates are more representative of the plume of substances that might enter groundwater. Elutriates can also fail to extract all toxicants. For example, root elongation in elutriates was a less sensitive indicator of on-site toxicity than seedling emergence in whole soil (Baud-Gasset *et al.*, 1993).

Solvent extracts are sometimes used to estimate the total quantity of toxic contaminants in a soil sample, because many particle-bound chemicals appear to be available to plant roots and soil organisms (Donnelly *et al.*, 1991). Different solvents extract different quantities and proportions of contaminants, but it remains unclear which components are bioavailable. Results of toxicity tests with solvent extracts should be interpreted with caution until better correlations are established with the toxicity of whole soils and environmental effects. Tests with whole soil should be regarded as the best indication of environmental hazard.

4.13 *Toxicity of Substances*

Key Guidance

- *Most substances can be tested in a straightforward fashion by means of the standard methods described by Environment Canada and others.*
- *Some individual chemicals present practical difficulties in testing and in interpreting the significance of results. Chemical properties causing difficulties in aquatic tests include low water solubility, volatility, instability, sorption, rapid bioaccumulation, and colour.*
- *To interpret toxicity tests with substances which have one of the troublesome characteristics, first steps should be to determine whether the test method was adequate for dealing with the problem, the likely influence on the test result, and the reliability of results. A further step would be to evaluate how the troublesome property would affect real environmental exposures.*

Some substances have characteristics which make them difficult to test. An investigator should be aware of the influences of these characteristics on test results, and some discussion follows. An instructive and much more complete treatment of a class of chemicals that is difficult to evaluate is provided in the book on polymers by Hamilton and Sutcliffe (1997). This ubiquitous category of substances finds use in agriculture, food processing, mining, transportation, and wastewater treatment, to name just a few areas. The ecological assessment of polymers can be difficult since many of them are sparingly soluble or minimally reactive, and chapters of the book deal with environmental aspects and with specific subclasses of these complex substances.

4.13.1 Testing Substances with Troublesome Properties

Properties that cause difficulties in testing would include low water solubility, volatility, instability, tendency for sorption, rapid bioaccumulation, and colour (Whitehouse and Mallet, 1993).

If toxicity is beyond the solubility limit, the compound might simply be declared nontoxic at that solubility limit. Alternatively, a hydrophobic chemical's solubility might be enhanced by dissolving it in a nonpolar solvent (e.g., methanol, acetone) and mixing that with the test water. Solvents for different classes of compounds are recommended in standard test-method documents. A solvent control is used to account for any toxicity from the solvent alone, but the entire approach is somewhat less than satisfactory. The solvent might make a toxic contribution that would be too small to show up in the control, but would nevertheless combine with the toxic contribution of the chemical to distort the results of the test. Toxicity of hydrophobic compounds can change greatly in the presence of a solvent, and can also vary with the solvent (Calleja and Persoone, 1993). Cautious interpretation of endpoints is in order for tests with solvents.

An alternative to solvents, when testing hydrophobic but liquid chemicals in water, is to attempt to disperse the test substance through the water as an emulsion. This is an especially suitable approach for oily substances. Testing petroleum hydrocarbons and their water-accommodated fractions is an important and complex topic. Test procedures are reviewed in Girling *et al.* (1992) and more recently by Environment Canada personnel (Blenkinsopp *et al.*, 1996). Unstable emulsions require continuous mixing to prevent droplets from coalescing, a particularly challenging task in tests with small organisms like *Daphnia*. Identically mixed controls must be set up. Assessing results from such a test would require detailed information on procedures, with thorough and representative chemical measurements, to demonstrate how well the emulsion persisted.

The toxicity of volatile substances can easily be underestimated if the concentration in test vessels diminishes quickly. Calculations must be made in terms of concentrations actually measured in the test tanks (always a good practice) rather than using nominal concentrations. Sealed systems with minimal headspace are recommended, but technically difficult to achieve, particularly in long exposures. A good procedure would be flow-through tests with a high turnover rate, if that could be accomplished for the toxic substance of interest. Volatile compounds are also a problem for soil testing, and suitable methods are still under development.

Chemical instability produces effects similar to those of volatility (underestimating toxicity, overestimating endpoints). There is really no remedy for this problem that does not compromise the validity of the test. For example, unsaturated fatty acids are acutely lethal to fish at high concentrations, but exact LC50s cannot be determined in static tests because the acids decompose rapidly (Leach and Thakore, 1977).

Some other categories of chemicals have parallel problems. Hydrophobic compounds sorb to surfaces such as those of the test vessels, reducing the concentration to which organisms are exposed. Rapid bioaccumulation produces a comparable effect. Tests with any such substance must use concentrations actually measured in the test water at appropriate intervals. Glass containers usually show the least tendency to sorb toxicants.

Colour or turbidity could be a practical consideration in tests like the bacterial luminescence test, and in algal growth tests that measure light production or absorbance. Corrections from the use of controls is sometimes attempted. Full details of measurements in test solutions would be necessary to interpret such tests.

The first step in interpreting tests involving any of these problems would be to determine whether the procedures took proper account of the troublesome property, and to what extent results were influenced and could be considered reliable. The second step would be to evaluate how the property would affect exposures in the actual receiving environment. Spills of a highly volatile substance, for example, might not produce a long enough exposure to cause observable toxicity. A constant discharge of the same volatile substance might achieve a hazardous level, however, especially if it were released under ice. Hydrophobic substances might partition rapidly out of the water, but could pose a long-term hazard in sediments, particularly if they were persistent and bioaccumulative.

4.14 Toxicity of Surface Waters

Key Guidance

- *Measuring the toxicity of surface-water samples, or testing in the receiving environment itself, is a distinct and useful application of testing. (Tests with sediments and soils are largely designed to serve this function, as covered in other sections herein.)*
- *Tests with surface water could supplement or parallel other measurements in a weight-of-evidence approach to pollution investigations. The tests could be correlated with measurements of effluent toxicity, chemistry, and dilution, as well as biological surveys, to rate toxicity as a contributing factor.*
- *Testing toxicity of the receiving water is of particular use in demonstrating the effects (or lack of effects) of acutely toxic discharges or spills, pesticide spraying, etc.*
- *The concept of requiring absence of toxicity beyond a mixing zone has gained wider acceptance in the past decade.*
- *Tests with receiving water use normal procedures, as described in Environment Canada methods documents. However, tests done* in situ *can only be single-concentration and they are subject to fluctuating concentrations of the toxic substance. Testing surface-water samples in the laboratory avoids those limitations.*

Measuring toxicity of surface waters, whether in the laboratory or in the actual receiving water, represents a discrete category of testing which can be useful in many circumstances. It can provide a definitive assessment of the toxicological status of surface waters. Tests can be done in the laboratory with samples collected from specific locations, or caged organisms can be exposed in selected locations (*in-situ* tests). [In tests with sediment and soil, such assessment of the receiving environment is the usual purpose; samples are collected in the field and then tested in the laboratory (see Sections 4.11, 4.12)].

The techniques for carrying out tests with surface waters and applying the results are not particularly unusual. Standard Environment Canada aquatic test methods can be used, and most of them contain specific instructions for assessing receiving waters (EC, 1990a,b,c; EC, 1992a,b,c,d,f; EC, 1998). Sampling methods, exposure techniques, and the role of ambient and *in situ* toxicity tests in Canadian environmental monitoring have been described in a technical guidance document (EC, 1997e). That document provides broad and useful advice, including methods of sampling effluent plumes in estuaries and coastal waters.

Tests of surface water could be used on their own, or with other techniques in a "weight-of-evidence" approach. They could be used in an exploratory mode, to see if a problem existed and should be further investigated, or they could be used in a confirmatory mode to check whether problems predicted from tests of toxicity or chemistry in the laboratory, actually existed in nature. In that respect, toxicity tests with surface water would serve a function that somewhat paralleled the use of biological surveys of resident communities. The community survey would be the more powerful tool, since it is focused on a higher level of biological integration, but the toxicity tests would usually be a more economical tool.

Toxicity tests on surface water could be used in conjunction with other techniques, for particular purposes. The manner of designing such parallel programs and interpreting the results is, to a large extent, self-evident. Some of the combinations of techniques follow.

- *Relating results of toxicity tests with surface water (or* in situ) *to estimated concentration of an effluent or chemical at a series of locations.* This is one of the most direct and convincing ways of demonstrating whether or not an effluent or chemical has deleterious environmental effects. For example, it is much more straightforward, and probably more convincing, than the predictive method of the USEPA using calculations with toxic units (see Section 4.10.2).
- *Checking whether modifying factors in surface waters changed the toxicity of an effluent or chemical.* The surface waters could detoxify by binding to organic molecules, or could enhance toxicity by changes in pH, hardness, etc. An investigator might attempt to predict such modifications from known relationships, or by tests in the laboratory which mimicked the modifying conditions, or used "upstream" water. However, testing samples of the receiving water would provide a simple, direct, and definitive answer.
- *Relating receiving-water toxicity to effects seen in resident communities.* This could determine whether the effects were accounted for by toxicity or whether there was an apparent fore of other factors such as habitat degradation.

4.14.1 Laboratory Tests of Receiving Water

To evaluate the toxicity of an area, samples of surface water can be collected from locations where toxicity is expected, and tested in the laboratory. Lethality in surface waters is now uncommon, but testing the surface water is a clear-cut way of assessing the seriousness of spills or pesticide spraying programs. To assess the generally lower modern-day levels of pollution, tests could make use of the newer fast sublethal tests. Sublethal tests with invertebrates are especially convenient if a fast measurement of sublethal effect is desired, or if a small volume of sample is desirable to reduce the cost of shipping to the laboratory.

Testing samples in the laboratory has some major advantages over *in-situ* tests (e.g., caged fish). Standard conditions of temperature and similar features can be maintained. Different concentrations of the surface water can be tested if desired, yielding a superior endpoint based on concentration rather than time for an effect. Cost of shipping samples to the laboratory could be balanced by an economy of effort in running standard tests in the laboratory, and by absence of additional travel costs to check caged organisms in the field.

Sublethal "ambient" tests were one of the tools used by USEPA in its field studies for validation of laboratory toxicity tests (Section 4.1.1). Most of these were done in the laboratory, using samples of surface water. Guidance on testing surface waters for U.S. purposes is provided in USEPA, 1991a.

4.14.2 **In-situ** *Tests*

Caged fish have traditionally been used to demonstrate the presence of acute lethal toxicity from discharges. A strong justification for such tests is that they are very convincing if damaging effects are shown among the tested organisms. The tests also cumulate all variations in magnitude of pollution; a peak of discharge would not be missed by a sample taken at only one (other) time.

A good example of *in-situ* toxicity testing is a study that was done in Prince Edward Island on treatment of potato fields with a fungicide (Ernst *et al.*, 1991). Various fish and native invertebrates were held in mesh cages in ponds exposed to spraying. Appreciable mortality of sticklebacks and water boatmen occurred. The tests were part of an integrated study that included water chemistry, biological surveys, and toxicity tests in the laboratory on the fungicide itself. The effects on caged organisms provided convincing evidence that spraying had some aquatic side effects.

Caged bivalve molluscs have often been used as *in-situ* monitors of the pollutional status of waterbodies, and an example of good procedures is provided by Salazar and Salazar (1996). The bivalves can be left in a location for periods of weeks or months, with no need for feeding or other care. Measured bioaccumulation of contaminants can provide excellent assessment of the general degree and kind of pollution over the exposure period. Survival and growth can also be assessed, i.e., a sublethal toxicity test.

There are two special considerations for interpreting the results of toxicity tests *in situ*. These do not necessarily apply to laboratory tests on samples of surface waters. First, the test involves only one concentration, i.e., at the full strength of the surface water. Results of the test could be expressed as the time required to cause a given effect, as a percent effect, or as a qualitative statement that the sample was, or was not, toxic.

The second important consideration is that *in-situ* exposures in surface waters would likely involve fluctuating concentration of the toxic substance(s) and of modifying conditions. It would be desirable to know the degree of variation, and that might be amenable to measurement by chemical analysis, if the toxicants were known or if some feature of the effluent such as conductivity could be defined.

4.14.3 Sublethality Beyond a Mixing Zone

Overt or implied requirements for absence of sublethal toxicity in the receiving water can be seen in recent approaches to regulating effluent discharges. In Quebec, since 1980, there has been a requirement for no sublethal effects beyond a mixing zone. Section 5.2.7 outlines Quebec's clearly defined mixing zones for that purpose. Distinct regulatory requirements for absence of sublethal effects outside a mixing zone have been found in the USA since the mid-1980s (USEPA, 1984a; 1991a). Alberta has also established requirements for stipulated mixing zones, beyond which the water should meet chemical guidelines intended to protect from sublethal effects during chronic exposures (AEP, 1995; Section 5.2.7).

In the Canadian requirements for monitoring environmental effects (EEM) of pulp mill discharges, there is implicit incorporation of the concept. One way of defining the effluent plume in the surface water is by the isopleth at which toxicity decreases to 0.1 IC25 in a series of defined toxicity tests (EC, 1997c; 1997e). Because that is such a low level of toxicity, the isopleth cannot be located by applying the tests to samples of the actual surface waters. It would have to be done by testing the effluent and calculating for dilutions as determined by field survey. A specific application of defining the plume in terms of sublethal toxicity is described in Section 5.2.4.

4.15 Conditions that Modify Toxicity

Key Guidance

- *Ambient conditions and biotic factors can modify the effects of toxicants in nature. These modifiers must be taken into account when extrapolating results of toxicity tests to the field.*
- *The major physicochemical entities influencing toxicity in aquatic environments are temperature, pH, light, hardness, salinity, dissolved and suspended organic matter, and dissolved oxygen.*
- *For sediments, particle-size distribution, porosity, redox state, and the physical nature of clays and organic matter are important variables, in addition to most of those mentioned for water.*
- *In soils, special modifying factors are type of soil, type of clay, particle size distribution, moisture content and drainage, organic matter content and type, cation exchange capacity, and natural toxicants.*
- *Important biotic modifiers include organism behaviour, life cycles, body size, nutrition, adaptation, interactions between species and the possibility of multiple stressors.*

The toxicity of substances, wastewaters, sediments, or soils can be modified by various environmental conditions, and by the features of the organisms themselves. In laboratory toxicity tests, these modifying factors are controlled and standardized, usually at favourable levels for the test organisms. However, in nature the levels are often less ideal and more variable. How the modifying factors are likely to change toxicity must be considered, therefore, when extrapolating test results to the natural world.

The modifiers can be broadly categorized either as physicochemical conditions (temperature, hardness, etc.), or biotic characteristics (e.g., species, age, behaviour, interactions among species). Most of the information on modifiers is specific to aquatic toxicity. The general principles also apply to modifications of soil and sediment toxicity, and some special items for those semi-solid habitats are described. In the USA, differences in results of aquatic toxicity tests caused by abiotic modifiers in field tests compared to the findings of laboratory tests, are described as the "water-effect ratio", an example of jargon which would not seem necessary.

4.15.1 Abiotic Modifiers

Key Guidance

- *Higher temperature often increases toxicity, but that is far from universal. Light can degrade toxicants or, for some PAHs, stimulate higher potency.*

- *Hydrogen-ion concentration is probably the most important physicochemical modifier, especially for metals, but also for ammonia and other ionizable substances for which the nonpolar and undissociated form is more toxic. Oxidation-reduction potential is important in sediments, where toxic forms of substances often prevail under reduced conditions (anoxia).*

- *Hard water decreases the toxicity of most metals. Salinity has little general effect on toxicity, for organisms living in their accustomed salinity. Organic matter can sorb and reduce the toxicity of some substances, whether the organic matter is dissolved, suspended, or a component of sediment or soil.*

- *Moisture content is an important feature governing processes in soils. Low dissolved oxygen exacerbates toxicity in water and sediment.*

Temperature. It might be expected that higher ambient temperature would increase toxicity, and that an effect would be produced by a lesser concentration or by a given concentration in a shorter time (Roch and Maly, 1979). That might be expected because the rates of chemical reactions between toxicants and cell components increase at higher temperature. In cold-blooded organisms, metabolic rates increase with temperature, tending to increase the rate of transport of chemicals across cell membranes. Higher metabolic rate creates a greater demand for oxygen, and in fish that would mean more water passing over the gills, and consequently, an augmented uptake rate for toxicants.

However, greater toxicity at higher temperature is not universal, and in fact there are many examples of toxicity driven both upwards and downwards at higher temperatures. One reviewer generalized that there can be major effects of temperature on short-term lethality of toxicants, but at longer-term lethal and sublethal thresholds, effect of temperature was generally less than two-fold, and could go in either direction (Sprague, 1985).

There are diverse findings in the literature on whether accumulation of metals from sediment speeds up or slows down with an increase in temperature. Metabolic adjustments can partly compensate for moderate temperature changes, and if chemicals are cleared from the body (depuration) more quickly than they enter, then toxicity can decrease at warmer temperatures (Dixon and Hilton, 1985; Green *et al.*, 1988; McGeachy and Dixon, 1990).

In soils, the highest (summer), lowest (winter), and most variable temperatures occur at the surface. At 20 to 30 cm of depth there are only slight daily cycles, and beyond that, seasonal variation decreases and temperature comes closer to the mean yearly air temperature.

Temperature can affect a number of chemical processes in water. It can change ionizable compounds from toxic undissociated forms to less toxic ionic forms. The effect tends to be smaller than the influence of pH, but is noticeable for ammonia, cyanide, and some organic acids and bases. At higher temperatures, the capacity of water to dissolve oxygen and carbon dioxide is substantially lower. Fresh water at sea level can hold >14 mg/L of dissolved oxygen at 0°C, but <8 mg/L at 30°C. Volatile compounds leave solution and pass into the atmosphere more rapidly as temperature increases.

Hydrogen-ion concentration. The pH of a solution is probably the single most important physicochemical factor modifying toxicity. It is especially important for dissolved ionizable compounds, since pH directly influences their dissociation and ionic state. Nonpolar organic compounds tend to accumulate in lipids and organic tissues, consequently they are much more toxic than the more soluble ionic forms.

Ammonia is an extreme example of a substance that changes its toxicity with pH. It is present in solution as un-ionized ammonia (NH_3) or the ammonium ion (NH_4^+). As shown in Table 6, at neutral pH virtually all of the chemical is present as relatively low-toxicity ionized ammonium, but at pH 10 and high temperature, it becomes mostly un-ionized ammonia, which has considerably more toxicity. Toxicity can be estimated from chemical measurements using information in Alabaster and Lloyd (1980) or Russo (1985).

The same pH-dependence of toxicity extends to other compounds that ionize. Bases (e.g., organic bases) become more toxic in alkaline water. Acids such as resin acids, hydrogen cyanide, and hydrogen sulphide become more toxic in acidic water. The strength of the pH effect varies with the tendency of each particular chemical to dissociate. Some highly divergent effects of pH on toxicity of different pesticides are shown by Doe *et al.* (1988).

The effect of pH on toxicity of metals is frequently very complicated because metals can show numerous chemical forms. As a broad generalization, dissolved metal ions are the main toxic species, and become more biologically active as pH increases (Cusimano *et al.*, 1986; Peterson *et al.*, 1984). However, overall toxicity usually decreases greatly in alkaline water, because the metals form relatively insoluble complexes with oxygen, hydroxide (OH^-) or carbonate (CO_3^{2-}), and become much less biologically available and less toxic (Howarth and Sprague, 1978; Stumm and Morgan, 1981; Bradley and Sprague, 1985; reviewed in Campbell and Stokes, 1985).

Differences in pH result in some of the major difficulties in interpreting laboratory toxicity tests. Many routinely tested effluents contain pH-sensitive toxicants, and are chemically unstable so that pH might change during a 4-d toxicity test or between renewals of solution. Active aeration of the test solutions could increase the pH by stripping out dissolved carbon dioxide, a weak acid (Abernethy and Westlake, 1991). Toxicity could change, and if the test pH differed appreciably from that of the receiving water, the findings could be of limited relevance.

Some effluents can be very acidic or alkaline, and current Canadian practice in definitive regulatory tests is not to adjust the pH (EC, 1990e; f). The test pH might be quite different from the receiving water, or even unfavourable for the test organism. Once discharged and mixed, the pH of the receiving water would prevail unless the effluent composed a very large proportion of the volume or flow. The implications of such differences in pH are currently a matter for discussion among scientists and regulatory personnel, but a person interpreting test results should always consider the ramifications of pH differences between the laboratory and the field.

In sediments, the pH can modify toxicity in a special way. Since pH largely controls the dissociation of polar compounds, including many organic compounds (phenols), it can markedly influence their tendencies to bind with organic matter in solution or in the solid phase, and that reduces bioavailability dramatically. Cationic metals are released from the solid phase into porewater as pH falls, especially at pH <7. Ammonia is common in sediments and in liquids derived from sediments, and often contributes to toxicity; the modifying action of pH, also applies in sediment tests.

In soils, there are similar effects in binding of polar substances and also nutrients, hence pH modifies their mobility and bioavailability. Most surface soils are slightly to very acidic because of organic acids. Soil pH is normally measured in a water-soil slurry and that obscures the fine gradients of pH in the interstitial water around soil particles. Accordingly, the kinetics of toxicants can be rather different than expected from whole-soil pH, and the latter should only be used as a broad indicator of the reaction of the soil.

Oxidation-reduction potential. This characteristic is usually referred to as redox potential, and is a critical modifier of toxicity in a sediment. It describes the degree to which oxidative or reducing chemical reactions predominate.

Table 6 Percentage of Ammonia Un-ionized in Distilled Water (APHA *et al.*, 1995)

Temperature (° C)	Percentage Un-ionized at Specified pH								
	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
5	0.01	0.04	0.11	0.40	1.1	3.6	10	27	54
10	0.02	0.06	0.18	0.57	1.8	5.4	15	36	64
15	0.03	0.08	0.26	0.83	2.6	7.7	21	45	72
20	0.04	0.12	0.37	1.2	3.7	11	28	55	80
25	0.05	0.17	0.51	1.7	5.1	14	35	63	84
30	0.07	0.23	0.70	2.3	7.0	19	43	70	88

In well-oxygenated locations (Eh positive), organic decomposition consumes oxygen and complete decay of organic compounds is possible. Reduced chemicals such as ammonia and ferrous iron (Fe^{2+}) are quickly oxidized to nitrate and insoluble ferric iron (Fe^{3+}). In anoxic environments (Eh negative), organic decomposition is incomplete and often produces acidic by-products, nitrogen remains reduced as soluble ammonia, and reduced metals accumulate. If oxygen returns, metals are removed from solution by co-precipitation with iron and manganese hydroxides (Krantzberg and Boyd, 1992).

Undisturbed sediments typically exhibit a steep gradient of Eh, from oxidized at the surface to increasingly reduced with depth. This results from the slow diffusion of oxygen into the sediments, especially if they are fine-grained or enriched (higher oxygen demand for decomposition). Benthic organisms can weaken the gradient by burrowing. These gradients are destroyed by most sampling methods and mixing sediments in the laboratory to homogenize them creates equal redox levels throughout the sample (Burton, 1991).

Light. Many biological effects modulated by light are important in toxicity tests. Many behavioural and physiological activities are influenced by light intensity or wavelength, and accordingly, measurements of such responses in toxicity tests are influenced by lighting conditions during the test or preceding it (Buikema and Benfield, 1979). Vertical migration, burrowing, or feeding can use light as cues, perhaps bringing organisms into or out of the influence of toxicants. For plants, the importance of light for sustenance and control of growth cannot be overemphasized.

Many organic compounds (e.g., chlorinated phenols) are unstable in light (photolytic) and can decay through the physical process of photo-oxidation. Accordingly, intensity of illumination could be important in a toxicity test. Strong sunlight and clear, shallow water would favour photolysis, while deep, turbid or highly coloured water would reduce it.

Photo-activation, however, occurs for some organic compounds, notably PAHs (e.g., anthracene, fluoranthene, pyrene). Ultraviolet light increases their toxicity, in some cases by orders of magnitude, and might also affect toxicity of metallocyanide complexes. Clearly, the type of lighting in laboratory tests could be of major importance in assessing toxicity if such substances were present. The instructions for most standard toxicity tests specify the intensities and wavelengths of light, usually with the intent of imitating natural light. However, artificial light including broad-spectrum or "daylight" lamps, differ considerably from natural light in the ultraviolet range. That difference could be of considerable importance. If PAH toxicity is of concern in a given situation (e.g., oil spills), it would be appropriate to apply the findings of recent research on the effects of lighting, this being a rapidly changing field of aquatic and sediment toxicology [44].

---

[44] The polycyclic aromatic hydrocarbon (PAH) anthracene is not toxic within its solubility range (< 35 µg/L). When the exposure occurs in the presence of ultraviolet light, however, anthracene is very toxic, with a 96-h LC50 to bluegills of 8 µg/L (Oris and Giesy, 1985). Six of 12 common PAHs are

Hardness of fresh waters. Hardness is defined as the sum of calcium and magnesium concentrations, both expressed as calcium carbonate in mg/L (APHA *et al.*, 1995). (Historically, hardness referred more loosely to the concentration of all cations that precipitated sodium soaps.) In most fresh waters, total dissolved solids (TDS, the mass of solutes remaining after a filtered sample is evaporated to dryness) are highly correlated with hardness, because calcium and magnesium are major cations in the water.

Toxicity of metals is greatest in soft water and declines steeply as hardness increases. Insoluble carbonates form (see above, pH) and calcium and magnesium ions compete with metals at the surface membranes of organisms, reducing bioavailability of the metal. The effect of hardness can be as large as two orders of magnitude and applies to both lethal and sublethal effects. Metal concentrations that would be acutely lethal in soft water might, in hard water, cause only sublethal effects or no observed effects (Figure 8). Dissolved salts and hardness are high in most industrial effluents, domestic wastewaters, leachates from buried wastes, and groundwaters, and that could mask potential toxic effects from metals. When such a waste is mixed with softer water in a receiving waterbody, dilution of hardness would tend to raise toxicity, while dilution of the metals would tend to reduce toxicity. That balance should receive due consideration in interpreting test results.

The changes in aquatic toxicity of metals provide the most striking examples of how important abiotic factors can be. The modifying effect of hardness is joined by the modifications caused by pH, alkalinity, and organic binding agents in the water. Generally, the soluble metals are the ones most easily available to organisms, but there continues to be uncertainty about suitable ways to measure the soluble or available forms. This complicated situation brings great difficulties to formulating environmental criteria, and the most suitable basis for regulatory limits remains in doubt. All these scientific and regulatory conundrums were delineated by a group of experts, who described the state of knowledge and possible new approaches such as regulations which used such esoteric approaches as computer models of metal speciation and uptake at the gill surface in aquatic animals (Bergman and Dorward-King, 1997). The publication would be useful to anyone concerned with metal toxicity and the influence of abiotic factors.

Salinity. Salinity originally referred to measurements of the total mass of dissolved salts in a given mass of solution. Today it is usually measured empirically from standard relationships of density or conductivity. Results are unitless (APHA *et al.*, 1995) but most people think in terms of g/kg or "parts per thousand" (‰). The average salinity of the oceans is 35, although values for coastal seawater are often in the range 28 to 32 and those would be more representative of the salinity used in toxicity tests with marine organisms.

The toxicity test methods of Environment Canada that are based on a marine organism, are consistent in their treatment of salinity when testing a sample that is primarily fresh water. There is adjustment using hypersaline brine, to bring the test salinity to the accustomed level for the test organism. This policy is a suitable one, supported by the information that follows.

Marine organisms are about equally tolerant of pollutants as their freshwater counterparts, *when each is tested in its own environment*. For most toxic substances, salinity itself does not appear to change toxicity greatly; the important aspect is whether the organism is marine, freshwater, or estuarine, and whether it is tested in its preferred water. Accordingly, interpretations involving salinity should focus on whether the organisms are stressed from unsuitable salinity, or whether they are in the process of acclimating to a new salinity.

A specific example of similarity between marine and freshwater situations is the finding that pulp mill effluent was equally toxic to coho salmon in fresh water and seawater, if the pH was the same and fish were acclimated to the water (McLeay *et al.*, 1979). Klapow and Lewis (1979) did an extensive review of data for 16 pollutants and showed that marine and estuarine organisms had distributions of tolerance that were very similar to the distributions for freshwater organisms. They concluded that any salinity effects on toxicity were obscured by more important factors. The USEPA (1978) came to the same conclusion.

---

subject to photo-induced toxicity (Oris and Giesy, 1987), and the enhanced toxicity extends to several species of fish and invertebrates (Hall and Oris, 1991; McCloskey and Oris, 1991). Photo-induced toxicity is reduced by dissolved organic matter, which absorbs the ultraviolet radiation (Oris *et al.*, 1990).

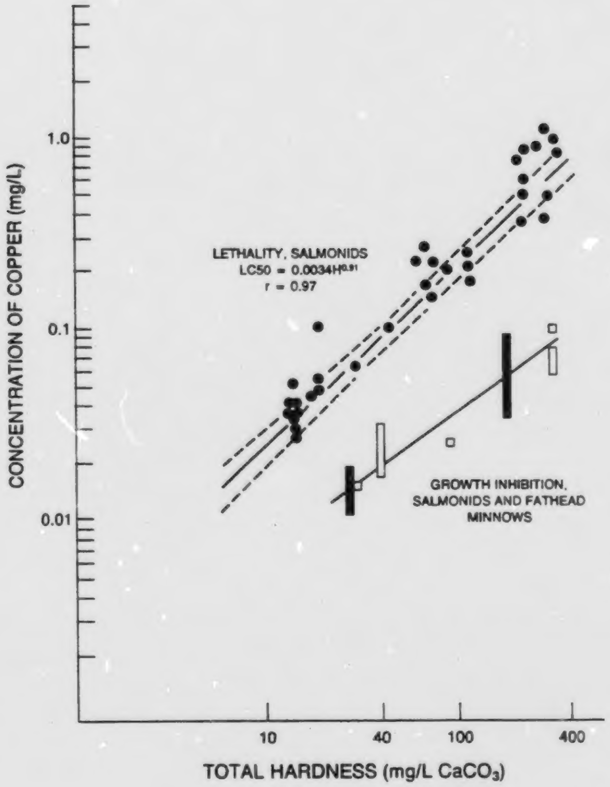

Figure 8 **Empirical Relation Between Copper Toxicity and Hardness of Fresh Water.** The upper line represents median lethality for most salmonids, the lower is for thresholds of growth inhibition for trout (open boxes) and minnows (dark) in alkaline water. From figures by Spear and Pierce (1979) and other data, modified by Sprague (1985).

Despite the similarities of toxicity mentioned previously, the numerical water quality guidelines for chemicals in fresh water (CCREM, 1987) are not designed for marine ecosystems and such use should not be attempted. Canada has interim guidelines for general variables of quality in marine and estuarine water (CCREM, 1987; Appendix 22), but these do not yet include toxic chemicals. There are also interim quality guidelines for marine and freshwater sediments (EC, 1995c), and British Columbia has quality objectives for water and sediment in the marine and estuarine reaches of Burrard Inlet.

In contrast, toxicity of pollutants can be expected to change when organisms move out of their native salinity. If a freshwater organism moves into water of increasing salinity, the general principle is that toxicity of most pollutants will decline to a minimum at the isosmotic point for the organism, usually about 1/3 of the full strength of seawater. Apparently an organism is osmotically "most comfortable" at such a salinity, and can better withstand toxic effects. At higher salinities, toxicity increases as might be expected from added salinity stress. Truly organisms appear to lose tolerance of toxicants as salinity decreases, presumably because of osmotic stress. Effects of toxicants on estuarine species would have to be interpreted specifically for the organism involved, based on what is known of its normal salinity range, and status of acclimation. Examples for these generalizations are given in Sprague (1985).

Organisms could easily be stressed in toxicity tests, from salinities beyond the favourable range. This could often be the case if a freshwater effluent were tested in various dilutions, using a organism. If salinities were not properly adjusted with brine, the higher concentrations of effluent would add increasingly greater stress from poor salinity, i.e., the response of the organisms could be partly the result of salinity, rather than purely of the toxicant(s). Such problems should be watched for when interpreting toxicity tests involving salinity changes. A good test would have measurements of salinity at each concentration (or some dependable estimates), and some comparison with the favourable range for the organism. In addition to total salinity, excessively high or low concentrations of the individual ions in seawater can have deleterious effects on mysids tested in the laboratory (Douglas and Horne, 1997).

For toxicity tests with marine and estuarine sediments, the same general comments apply. Exposing organisms to salinities outside their favourable range is poor procedure in a toxicity test. Salinity in tests can vary, depending on the technique adopted, for example, the technique used if there are dilutions of sediment or porewater. A less obvious but more likely problem would be differences in salinity of porewater within a sediment being tested, and the overlying water added to the test chambers. The porewater salinity would be influenced by (a) the field situation before collection, (b) manipulation of the sample in preparation for the test, and (c) ease of exchange during the test, between the porewater and the overlying water (because of particle size, porosity, etc.). For example, there could be a false positive effect in a toxicity test, because of unfavourable salinity in the porewater, if sediment from a freshwater or estuarine location were tested against a stenohaline, burrowing amphipod. An investigator must ensure that salinities are suitable for the test organism, and equitable in all phases of the exposure. Anyone interpreting the results of sediment tests should check whether proper salinity was used.

Dissolved organic matter and suspended matter. Natural organic matter can lower the toxicity of metals and hydrophobic substances by chemical and physical complexation, which reduces bioavailability. Most fresh waters contain about 0.5–25 mg/L of dissolved organic matter (DOM, which is also measured as dissolved organic carbon, DOC). Even relatively uncoloured waters in the middle part of that range can substantially reduce the toxicity of metals and pesticides [45]. Wetlands and drainages from boggy areas can have DOM of 50–70 mg/L (Day, 1991; Ortego and Benson, 1992), consisting mostly of stable, polycyclic, humic and fulvic acids, which cause the distinctive tea-brown colour of many Canadian surface waters.

Suspended organic matter also effectively binds metals and organic compounds and in high concentrations, can significantly reduce toxicity (e.g., St. Lawrence River estuary; Gagnon *et al.*, 1992). (DOM is the more important component of organic matter in most surface waters.) Suspended *inorganic* particles can have various effects on bioavailability and toxicity, depending on the toxicant and the effect being considered. High concentrations of inorganic suspended solids can sorb hydrophobic toxicants and decrease their uptake and toxicity (Pratt *et al.*, 1990; Schrap and Opperhuizen, 1990). However, the suspended particles can also retain toxicants in the water column and make them more available to organisms such as filter-feeders. Often, disturbance of contaminated sediments increases uptake of contaminants by organisms in the water column (Seelye *et al.*, 1982). If suspended solids were elevated in the test water, that could be a direct additional stress on the organisms being tested.

Moisture content. This is an important factor affecting all processes in soils. Water is the primary route of toxicant uptake by plants and soil organisms, except for organisms like earthworms for which ingestion is important. Lack of moisture can be a source of stress to plants during droughts. Too much moisture can affect soil invertebrates which often migrate vertically to avoid unfavourable moisture. Earthworms can die in waterlogged soil.

Dissolved oxygen. Increased toxicity to fish with lowered dissolved oxygen (DO) has been repeatedly observed for a wide variety of substances such as ammonia (Lloyd, 1961; Thurston *et al.*, 1981), metals (Voyer *et al.*, 1975), pulp mill effluents (Alderdice and Brett, 1957; Hicks and

---

[45] Hydrophobic chemicals (with a high octanol–water partition coefficient, K_{ow}) accumulate in living tissue (Section 4.6) and also form stable complexes with the humic and fulvic acids of DOM. The complexed chemicals are more water-soluble than the free chemicals, and less bioavailable. High concentrations of DOM can dramatically attenuate the toxicity of many hydrophobic organic compounds and metals (Stackhouse and Benson, 1988; Gardner and Gunn, 1989; Day, 1991).

DeWitt, 1971), and natural plant toxins (Kulakkattolickal and Kramer, 1987). This effect applies to acute and chronic toxicity, and even a moderate decline in DO can substantially increase the effect of toxicants.

There are reciprocal interactions between toxicity and low-oxygen stress. Many effluents, particularly pulp mill wastewaters, can have a high biochemical oxygen demand (BOD) and can cause a local DO depression in the receiving waterbody. The effluent's real toxicity under these circumstances might be underestimated by standardized laboratory tests, especially if the samples were aerated during the test.

At the physiological level, exposure to toxicants typically increases the respiratory rate of fish, and the interaction is particularly strong in toxicants that interfere with gill function, such as zinc and the resin acids of pulp mill effluent (Hicks and DeWitt, 1971). One possible result is a feedback situation in which the toxicant impairs oxygen transport, that induces increased gill irrigation, and that in turn accelerates toxicant uptake as more contaminated water passes over the gills (Lloyd, 1961; Birtwell, 1989). Alternatively, gill damage could impair oxygen uptake to such an extent that sublethal or lethal effects were caused by the oxygen lack, wholly or in part.

When extrapolating results of toxicity tests to the environment, there are particular situations in which it is important to consider enhancement of toxicity by low DO. Oxygen is often low in the hypolimnion of stratified, productive lakes in summer, and in some ice-covered lakes and rivers in winter. Free oxygen can be low or absent in sediments and in layers of water near sediment. Shallow waterbodies with heavy growth of aquatic plants frequently suffer low dissolved oxygen at night.

4.15.2 Biotic Modifiers

Key Guidance

- *Behaviour of toxicant-exposed organisms can be important, especially avoidance reactions which are understudied but can be equally serious as physiological damage.*

- *Early life stages in plants and animals generally include the most sensitive stages, which is a major reason that they are used in toxicity tests. Aside from that, body size has an unexpectedly variable role in tolerance of toxicants.*

- *Poor-quality nutrition can increase toxic effects. Adaptation to toxicants generally has a modest effect of two- or three-fold increase in tolerance.*

- *Multiple sources of stress can combine in the organism to make it more sensitive to all sources, whether they be natural or anthropogenic. Stressors could include adverse natural conditions such as high temperature, human activities such as dredging, or the presence of toxic substances.*

Behaviour. Failure to consider avoidance reactions can occasionally be a major error in assessing environmental pollution. Animals in standard toxicity tests are confined in a fixed exposure, but in the field, many species of fish are free to swim away from effluent plumes, and many invertebrates can drift downstream in a river. The classic example of this is in the field validation by Geckler *et al.* (1976), in which standard laboratory tests failed to predict the total effect in the ecosystem because they ignored behavioural actions of fish [in the river, fish avoided a large, lightly polluted section of river and would not spawn there (Section 4.1.1)].

Fish, however, are sometimes attracted to effluent plumes. A common finding is that fish congregate in warm effluent during winter. Obviously, that behaviour might expose them to harmful concentrations of toxicants in the effluent or create a combination of stressors such as sublethal toxicants together with rapid temperature change when they left the plume. Fish are very sensitive to water chemistry, and might also be attracted to effluent plumes by chemical cues. If that happens, it is poorly documented, but laboratory studies have shown apparent attraction of crustaceans to petroleum products (Atema and Stein, 1974; Percy, 1977), and fish have shown attraction to low concentrations of pulp mill effluent but avoidance of high ones, and avoidance of low concentrations of chlorine, but attraction to a high one because of an apparent "physiological trap" (Sprague and Drury, 1969).

A case can be made for always carrying out two sets of toxicity evaluations in the laboratory because of these possible behavioural effects: (a) the standard tests for physiological damage; and b. parallel behavioural tests, particularly for avoidance reactions and attraction. Persons evaluating polluted locales should at least review available findings on behavioural reactions. The topic is understudied, and existing publications deal mostly with avoidance reactions of fish.

In addition, the normal daily, seasonal, or life-cycle movements of mobile organisms might produce an irregular exposure to a discharge, rather than the constant exposure in laboratory tests. For example, migratory species of fish might be absent from a polluted area during some seasons.

Invertebrates are generally sessile, but toxic exposures and effects can still influence, and be influenced by, their behaviour. Soft-bottom organisms might escape surges of toxic discharges by burrowing, but perhaps at a cost in other functions such as feeding. Conversely, some species which normally burrow are used in toxicity tests, and their unwillingness to burrow provides a sublethal measurement of effect, for the sediment being tested. Many benthic insects spend part of their life cycle out of the water and would not be affected by toxic substances during that season. In flowing waters, a peak of pollution can stimulate benthic invertebrates to release and drift downstream which would deplete the local population.

Life cycles. Sensitivity to toxicants varies with the life stage of the organism. Among most organisms, gametes and very young stages are the most sensitive, for example newly hatched larvae, newly germinated seeds, and reproductive stages of algae. This point is obvious, since environmental toxicology has taken advantage of the feature to develop short sublethal tests "with the most sensitive life stages". That underlines the importance of considering life stage in assessments of environmental pollution, as well as the need to protect reproductive phases. Crustaceans and aquatic insects are particularly sensitive immediately after moulting (Wajsbrot *et al.*, 1990), presumably because the new, uncalcified carapace permits an influx of water (Buikema and Benfield, 1979).

Variation of toxic effects with the time of year has also been observed, such as a seasonal pattern of sensitivity of zooplankton to metals (Borgmann *et al.*, 1980). Sometimes there are resistant stages such as cysts, that sustain an organism through unfavourable times and are often highly resistant to outside conditions including toxicants.

Body size. One might expect that small aquatic poikilotherms would be more sensitive to toxicants than larger ones. The greater ratio of surface area to volume in smaller organisms should allow relatively greater influx of toxicants per unit of body weight. This expected effect of body size is indeed often found among fish, and sometimes for invertebrates [cladocerans by Winner (1981) and Lazorchak and Waller (1993) and amphipods by Naylor *et al.* (1990)].

There are many exceptions to the above generalization, however, and among fish the effect of size might be described as chaotic. An example is copper tolerance, with sunfish showing the expected increase in tolerance with size, but trout showing no change, for tests in the same laboratory (Anderson and Spear, 1980). Trout in another laboratory, however, did show an increase with size (Howarth and Sprague, 1978). Trout from 1 to 50 g showed a direct relationship between size and increased tolerance of the insecticide permethrin, but from 50 to 200 g showed no further change (Kumaraguru and Beamish, 1981). Adult trout >1 kg are more sensitive to ammonia than any other size or life stage (USEPA, 1985a).

When interpreting toxicity findings, the possible effect of size should be considered.

Nutrition. The nutritional status of organisms can have a positive or a negative effect on toxicity. Hungry or starving animals, or plants suffering nutrient deficiencies, are already stressed and have less metabolic scope available to combat a toxicant than organisms with nutrient sufficiency (Dixon and Hilton, 1985; Holdway *et al.*, 1987; Cripe *et al.*, 1989).

Conversely, good nutrition sometimes enables fish to increase their tolerance of toxicants. Higher levels of protein in the diet resulted in a six-fold increase in LC50 for the insecticide chlordane (Mehrle *et al.*, 1977), and a two-fold increase for copper (Dixon and Hilton, 1981). Large amounts of vitamin C increased the chronic threshold of toxaphene up to four-fold (Mayer *et al.*, 1978).

Adaptation. Organisms exposed to sublethal concentrations of a toxicant for some days or weeks might adapt metabolically, and increase their tolerance of that toxicant. This phenomenon is called here, *adaptation*. Most observed increases have been of the order of a two-

or three-fold rise in LC50, and might not be of major importance in evaluating an environmental situation.

Adaptation has been observed repeatedly for fish exposed to metals (Dixon and Sprague, 1981a; Orr *et al.*, 1986; Hobson and Birge, 1989) or insecticides (Holdway *et al.*, 1987). Exposure to metals triggers an increase in metallothioneins within the body, which complex with metals and make them less damaging, an apparent defence mechanism (Dixon and Sprague, 1981b). Freshwater crustaceans from streams contaminated by zinc are less sensitive to zinc or low pH than similar organisms from unpolluted habitats (Naylor *et al.*, 1990).

No adaptation was found for trout exposed to lead, and they were instead rendered more sensitive to a second exposure (Tang and Garside, 1987). That might be expected since lead is a persistent bioaccumulative toxicant that is only slowly removed from the body. Adaptation to the herbicide atrazine has been demonstrated for some algal species but not others (Hersh and Crumpton, 1989). No acclimation was evident in fish exposed to parathion for five weeks (Banas and Sprague, 1986).

Multiple stressors. Exposure to sublethal concentrations of toxic substances creates a stress on organisms in the environment, similar to those created by low oxygen, temperature extremes, lack of food, predation, disease, or nutritional imbalance. These stressors have a cumulative effect on the populations. Fish in low-DO environments, for example, or nutrient-deficient plants or animals, are more susceptible to toxicants. Human exploitation such as harvesting of fish, can also create a stress on survivors which exaggerates the effect of toxicants (Barnthouse *et al.*, 1990). Conversely, exposure to sublethal concentrations of toxicants (toxic stress) can reduce an organism's capacity to withstand other stressors. Sublethal exposure to metal, for example, can reduce the thermal tolerance of fish (Carrier and Beitinger, 1988).

For sediments, small contributions of toxicants from multiple sources are particularly important since sediments can accumulate a complex mixture of hydrophobic contaminants (Chapman and Fink, 1984). Animals that burrow or ingest sediment are more directly exposed to sediment-borne contaminants than predators or species that filter food from the water column (Lamberson *et al.*, 1992). If the overlying waters are productive, periodic or semipermanent anoxia could be an additional stressor.

Soils polluted from industrial use also tend to contain a diversity of organic and inorganic toxicants. In addition, soils commonly contain natural entities that deter growth of plants. Many plants and fungi secrete inhibitory substances (allelopathic compounds) that impair the growth of competitors. The compounds are mostly substituted phenols and terpenes (Harborne, 1982). Other biocidal or insecticidal compounds can leach from fallen leaves and other plant litter and inhibit growth or metabolic function of microbes or soil invertebrates (i.e., non-target effects). Areas of the landscape with incomplete or unhealthy vegetation and inhibited soil processes might indicate toxicity problems, or the cause might be unsuitable physical conditions (e.g., erosion, compaction).

The cumulative impact of multiple environmental stressors including toxicants will ultimately determine the success of a population in nature. Thus, in extrapolating results of toxicity tests to the field, it is the combined influence of all environmental stressors that must be considered, including both natural and anthropogenic items such as fishing and dredging.

Applications of Toxicological Data

This section discusses the application of toxicological data to specific programs within Environment Canada. Many of the principles from Chapters 3 and 4 apply to the programs, but use of toxicity data can differ slightly to meet objectives of particular programs.

In this chapter, interpretive guidance is first given for the use of toxicity tests in programs which directly address the effects of waste discharges:

- regulatory toxicity tests on industrial effluents (Section 5.1);
- *Environmental Effects Monitoring* for industrial effluents (Section 5.2);
- disposal of dredging spoils and wastes at sea (Section 5.3); and
- Environmental assessments for new industries producing effluent (Section 5.4).

Then the use of toxicity tests in some more specialized programs of Environment Canada is summarized:

- assessment and remediation of *Contaminated Sites* (Section 5.5);
- methods for the development of environmental quality guidelines (Section 5.6);
- evaluating hazards of new chemicals under *CEPA* (Section 5.7);
- environmental assessment of *Priority Substances* (Section 5.8); and
- evaluating consumer products for the *Environmental Choice Program* (Section 5.9).

5.1 *Control of Effluents*

Key Guidance
- *Canada was one of the first countries to use toxicity tests on effluents as a regulatory tool.*
- *In 1992, new regulations for pulp and paper mills continued using lethality tests on the effluent as one of the major facets of the control program.*
- *The clear-cut nature of the toxicity results has made them useful tools for regulatory monitoring.*

Canada was an early user of lethality tests on effluents as a means of regulating them. In the 1970s, regulations and guidelines required that effluents from new pulp mills should be non-lethal to rainbow trout (EC, 1971; 1972). The federal regulations and guidelines outlined a simple toxicity test that had many good features for monitoring. The definitive test required ≥80% survival of rainbow trout in 65% concentration of effluent, which was thought to be state-of-the-art for wastewater treatment at that time. If fewer than 80% of the trout survived, the effluent was considered "toxic" and the discharge was an offence under Section 33 of the Canada Fisheries Act.

The regulatory approach for pulp mills defined an allowable limit of toxicity, was straightforward in procedure, and was moderately successful. A few years later, it was stated that "no other country demands that industry satisfies a biological criterion of acceptable effluent quality", and that "the *existence* of such standards has required newer [pulp and paper] mills to install secondary treatment facilities" (Howard, 1978). Some provinces, notably Alberta and Ontario, also used toxicity testing for less formal regulatory purposes, as early as the 1950s.

In 1992, new federal regulations for paper mills spurred major activities in bioassessment (DFO, 1992). Toxicity requirements and chemical limits were at the heart of the regulations, and the limits paralleled the performance of good, modern, secondary waste treatment. The mill effluents were required to be non-lethal to rainbow trout and *Daphnia* by the end of 1993. If mortality of trout was >50% in full-strength effluent during a four-day exposure, that represented a failure. Increased testing frequency was then required (weekly instead of monthly), and the LC50 had to be estimated in a multi-concentration test.

Tests with *Daphnia magna* were used for monitoring rather than as a definitive indication of toxicity. If mortality during a two-day exposure to effluent was >50%, this triggered more intensive testing. An immediate estimate of the LC50 for trout was required, as well as the LC50 for daphnids every three days, until three consecutive samples met the limits. A remedial plan was also required.

[The program also had requirements for a maximum biochemical oxygen demand of 7.5 kg/tonne of product, and maximum suspended solids of 11.25 kg/tonne. These chemical items are not discussed here. In addition, there was an extensive biological and chemical monitoring program for the receiving environment (see Section 5.2.)]

There were major improvements in liquid discharges from Canadian pulp and paper mills, during the years following introduction of the regulations in 1992. By the mid-1990s, almost every Canadian pulp mill had installed secondary waste treatment or had plans to do so. During the two years from 1994–95, a monitoring program for 52 mills with secondary treatment showed that 2/3 of them did not measure any episodes of effluent lethality (Kovacs and O'Connor, 1996). The program is serving as a general pattern for regulation of other industrial categories such as metal mining.

At the technical level, the strengths of the toxicity testing program for effluents were clear.

- relatively fast results
- relatively inexpensive
- definite results, easy to comprehend by the mills and by the public
- conducive to fast remedies if a problem developed
- encouraged progress towards a government goal of non-lethal effluents

The definitive and understandable nature of the results from an effluent toxicity test has made it a major tool for prosecuting offending industries.

The main technical deficiency of the effluent-testing program was the absence of any particular relationship between toxicity of an effluent and degree of damage to the receiving ecosystem (or lack of damage). The relationship would also depend on relative volumes of the effluent and the waterbody, rate of water exchange, rapidity of dispersion, and other variables. Those aspects were covered by the other part of the regulatory program which assessed environmental effects (Section 5.2).

5.2 Sublethal Toxicity Data in Environmental Effects Monitoring

Key Guidance

- *Regulations for environmental effects monitoring (EEM) by the pulp and paper and metal-mining industries require sublethal tests of the effluent with species of fish, invertebrates, and plants.*
- *There are no pass/fail limits, but results are used in combination with chemical measurements and surveys of the receiving water, to evaluate any potential effects, and the success of effluent management.*
- *Data from sublethal toxicity tests must be validated before they are used. There are specific requirements for validity and for reporting under an EEM program, in addition to the requirements of each standard test method.*
- *The sublethal tests can provide useful documentation of changes in quality of an effluent over time.*
- *Over a three-year period, an EEM program will generate four to six endpoints of sublethal toxicity for a given test species and effluent, a relatively small series for evaluating variability. A single value is selected to characterize the endpoints, for comparison with concentrations observed in the discharge plume in the receiving water. The single value for a given species and location may be selected by one of three methods.*

1. *The lowest IC25 (or EC25) of the series may be selected. This choice is only recommended if the series contains 10 or more endpoints.*
2. *The geometric mean of three or more IC25s may be used.*
3. *A statistical procedure calculates the Predicted Minimum Toxic Concentration (PMTC) of effluent, from three or more endpoints. Only 5% of samples would be expected to be more toxic than the PMTC, at the 95% level of confidence.*

- *The selected descriptor of sublethal toxicity may be superimposed on the observed concentration isopleths of a given plume, as a "zone of potential effect".*
- *The relative importance of multiple discharges can be evaluated by applying specific techniques to results of sublethal tests on the effluents and samples of the waterbody.*

The first formal Canadian program of Environmental Effects Monitoring (EEM) was applied under the Pulp and Paper Liquid Effluent Regulations of the Canada Fisheries Act (DFO, 1992). A second application of EEM is scheduled to commence during the year 2001, under the amended Metal Mining Effluent Regulations. Regulatory provisions that use EEM focus on assessing whether there is protection of fish populations, fish habitat, and use of the fisheries resource, in waterbodies which receive effluents. Results of EEM are interpreted by a "weight-of-evidence" approach using a suite of complementary tools, including sublethal toxicity tests on the effluent. Specific technical guidance has been developed for both the pulp and paper and metal mining EEM programs (EC, 2000b, c) and should be consulted by program participants.

The program for pulp and paper mills requires three kinds of sublethal tests on the final effluent, twice a year. Data are reported to government within 90 days of test completion, and are interpreted in combination with results of field monitoring, at the end of each three-year cycle (EC, 2000a). The proposed EEM for metal mining will be more site-specific and tiered, and will not operate on a fixed three-year cycle. Six sets of data from a battery of sublethal toxicity tests will be required during the first three-year period. After the initial three years, there will be an ongoing annual effluent monitoring requirement for sublethal toxicity testing under the mining EEM program. Depending on the likelihood of effects in the receiving water, and the mixing characteristics at the site, further sublethal toxicity testing on effluent or effluent-exposed surface water samples might be recommended by the mining company in a future EEM study design or required by the Regional Authorization Officer following discussions with the mine and other stakeholders.

Sublethal toxicity tests complement other components of EEM: (a) chemical measurements in the effluent, the receiving water, the sediment, and fish tissue; (b) cages in the receiving water, of cultured or indigenous organisms; and (c) biological surveys of organisms living in the waterbody (e.g., macrophytes, benthic invertebrates, fish). Within EEM, sublethal toxicity tests on effluent or receiving water can be used to:

1. measure changes in discharge quality over time, as a result of process changes or effluent treatment at the facility;
2. estimate the potential effects in the receiving water; and
3. estimate, in multiple discharge situations, the relative contributions from various point and/or non-point sources, to any observed effects in the receiving environment.

Canadian programs for EEM require sublethal toxicity tests on the final effluent, or on samples of receiving water. Typically, a battery of three or four tests is used: (1) early-life-stage development of fish; (2) reproduction of an invertebrate; and (3) growth of an aquatic plant. The choice of tests and species depends on whether the receiving water is fresh or estuarine/marine (Table 7).

5.2.1 Validation of Test Results

Regardless of the final use of results from sublethal toxicity tests, the staff at a regulated facility should first ensure that their testing laboratory has generated valid data and that the report contains all the required information on quality assurance. Procedures for quality assurance and quality control (QA/QC) must be followed by both the field crews collecting samples and the laboratory carrying out the testing, as discussed in the requirements for EEM, in technical guidance documents for the pulp and paper industry (EC, 2000a, b), and in the relevant methods documents for the toxicity tests. These measures optimize validity and precision within the limitations of the test method and natural biological variability.

Problems with QA/QC can lead to rejection of the test results and the need to collect and test a new sample. That could happen whether the problem was identified by staff of the testing facility, or by government regulators after the data had been submitted. An industry/government Expert Working Group on Laboratory Toxicology reviewed the data on sublethal toxicity from Cycle One of

Table 7 Acceptable Sublethal Toxicity Tests for Environmental Effects Monitoring

Description of Test	Receiving Environment	Test Species	Acceptable Method(s)
Early life-stage or development of fish	Marine/estuarine	Inland silverside or topsmelt	USEPA, 1994a USEPA, 1995
	Fresh water	Fathead minnow or salmonid*	EC, 1992b or EC, 1998
Reproduction of an invertebrate	Marine/estuarine	Echinoid (sea urchin or sand dollar)	EC, 1992f
	Fresh water	*Ceriodaphnia dubia* (waterflea)	EC, 1992a
Toxicity to an aquatic plant	Marine/estuarine	*Champia parvula* (macro-alga)	USEPA, 1994a
	Fresh water	*Selenastrum capricornutum* (alga) or *Lemna minor* (duckweed)	EC, 1992d or EC, 1999

* For locations west of the Rocky Mountains, where fathead minnows are not native, a salmonid species must be used according to EC (1998).

the pulp and paper program for environmental monitoring. Because numerous problems of quality were identified, the working group recommended that the revised requirements (EC, 2000a, 1997c) should incorporate a number of specific QA/QC practices, including the following ones.

- A sublethal test must be repeated on a new sample if the test fails to meet the criteria for validity which are outlined in the document describing the test method.
- A sample of effluent must be tested within three days of collection, or else re-sampling is required.
- Reporting must meet the minimum level required by the test method.
- Results must be reported within 90 days of completing a test, to allow a check of quality assurance.
- Reporting of "less than" values as a test endpoint is not acceptable.
- At least one test concentration must have been higher or lower than a reported endpoint.
- A test with a reference toxicant must be available for a date that is within 30 days of the effluent test.
- Tests with a reference toxicant must be performed under the same experimental conditions as the effluent test.

These requirements for QA/QC are intended to ensure that valid and high-quality data on toxicity will be generated by private testing laboratories during Cycle Two and future cycles of the pulp and paper program.

Similarly, data could be declared invalid if one or more essential elements of a test method were not present, e.g., failure to meet criteria for health of organisms, inappropriate manipulations of the sample, failure to conduct the required monitoring during the test, or use of incorrect procedures for calculating the endpoint.

Toxicity data submitted as part of the EEM program for the pulp and paper industry must be accompanied by description of the materials, methods, and calculations for each test. Minimum reporting requirements are detailed in Section 8 or 9 of each method document for a toxicity test issued by Environment Canada. An example of minimum reporting requirements for the *Ceriodaphnia* test is outlined in Table 8. Although general reporting requirements for the USEPA methods (1991b and 1995) are given in Section 10 of the EPA documents, Environment Canada has established minimum reporting requirements for these methods when used in EEM. Checklists have been prepared, which summarize minimum requirements for reporting components of quality assurance and test methods. The checklists are intended to assist both staff at regulated facilities and government regulators, in assessing of test information, and are available from Environment Canada via the Internet at the web-site for EEM (www.ec.gc.ca/eem).

It is recommended that laboratories should be accredited if they are to be contracted for sublethal testing by the pulp and paper or metal mining industries. Accreditation can be done by the Standards Council of Canada through the Canadian Association for Environmental Analytical Laboratories (CAEAL), by the Ministère de l'Environnement du Québec (MEQ), or by an organization offering an equivalent level of accreditation. These organizations follow accreditation standards of the International Standards Organization (ISO). Such a level of quality assurance should be a prerequisite for selecting a testing laboratory (EC, 2000a; 1997c).

5.2.2 Tabulation of Sublethal Toxicity Test Endpoints

Valid test results should be available for all test species (plants, invertebrates, and/or fish) for each of the sampling periods in a cycle or tier of EEM. The IC25 (or EC25 for the salmonid embryo test) should be extracted from each test report. If the effluent does not cause a 25% sublethal inhibition or effect in a freshwater sublethal test, then an IC25/EC25 cannot be calculated and the endpoint should be reported as >100% concentration. For estuarine/marine tests, the highest possible concentration will be 67% effluent because of adding hypersaline brine, and an IC25 of >67% would be reported for tests showing <25% inhibition at the maximum concentration.

Mortality in some of the concentrations might also prevent calculation of the IC25/EC25. For example, there might be no measured effects (mortality, growth or reproduction) in 32% concentration, but complete mortality in 56%. It would then be impossible to obtain an estimate of the inhibition of growth or reproduction for the 56% concentration, and hence difficult to calculate an IC25/EC25. In such a situation, if an IC25/EC25 cannot be calculated, then the result should be assumed to be greater than the highest concentration that caused no effect, in this case "greater than 32%".

Once validation and tabulation of the sublethal toxicity data have been completed, the next step is interpretation. Guidance on applying and interpreting the results is provided in Sections 5.2.3 to 5.2.5, which deal with the three principal uses for sublethal data within an EEM program.

5.2.3 Monitoring Temporal Changes in Effluent Quality

Sublethal toxicity endpoints can be used to compare the quality of effluent at different times. The comparison might be between EEM cycles or repeated monitoring efforts at a given site, or to assess trends at similar types of facilities. Comparisons could be based on individual IC25s, or on geometric means of the IC25s *(GM-IC25* or *GM-EC25)* for each species. If a geometric mean of IC25s is calculated, any tests which do not cause 25% sublethal inhibition should be assigned an endpoint of 100% for purposes of the calculation[46].

The effectiveness of sublethal toxicity tests in identifying a trend can be seen in data from Cycle One of EEM for the pulp and paper industry. Between 1992 and 1996, many Canadian mills were upgrading or building effluent treatment systems. As the new or improved treatment systems became operational, there were many cases of decreases in sublethal toxicity of effluents (EC, 1997c). When results for all the Canadian mill effluents receiving secondary treatment were separated from data of mills using only primary treatment, there was clear evidence of reduced toxicity when secondary treatment was in place or recently implemented. That difference was seen for sublethal tests with *Ceriodaphnia*, *Selenastrum*, fathead minnows, and inland silversides. The ability of the

---

[46] Although this action might seem to be biased against the industry being evaluated, that is really not the case. Assigning a value of 100% will almost always mean that this hypothetical IC25 is closer to the other measured IC25s, reducing the standard deviation.

Table 8 Minimum Reporting under EEM, for Sublethal Tests with *Ceriodaphnia dubia*

Effluent

- sample type (e.g., whole effluent, final effluent)
- information on labelling/coding of sample
- temperature of sample upon arrival at laboratory
- date for sample collection; date and time of receipt at laboratory

Test Organisms

- species; source
- age at start of test
- unusual appearance, behaviour, or treatment of test organisms before their use in the test

Health Criteria

- mean percent mortality of brood organisms in individual cultures during 7 d preceding test
- mean number of surviving young per adult, in individual cultures during 7 d preceding test
- number of young produced by each brood organism in its previous brood
- ephippia in the culture

Test Conditions and Facilities

- test method
- dates or days during test when subsamples or multiple samples were used
- dates for start and completion of test
- description of test vessels (size, shape, type of material)
- person performing test
- rate of pre-aeration of solutions, if required, before adding organisms to the test vessel
- duration of pre-aeration
- rate and duration of aeration, if any, during exposure of test organisms
- procedure, if any, for adjusting pH of sample
- procedure, if any, for filtering and adjusting hardness of sample
- type(s) and source(s) of control/dilution water
- type and quantity of any chemical(s) added to control/dilution water
- concentrations of test solutions
- volume and depth of solution in each test vessel
- number of replicates per concentration
- number of organisms per replicate
- dissolved oxygen and pH of sample just before its use
- temperature, dissolved oxygen, and pH of test solutions and controls
- anything unusual in the test, deviation from method, problem, remedial measure

Results

- mean percent mortality of first-generation daphnids for each test solution and observation period
- mean number of neonates per first-generation daphnid for each test solution and observation period
- LC50 (and 95% confidence limits); indication of quantal statistic employed
- IC25 (and 95% confidence limits) for reproduction; indication of quantitative statistic used
- results and duration of reference toxicant test(s)

---

sublethal tests to detect improvements in effluent quality was important during the review of data from Cycle One in the EEM program (EC, 1997c).

5.2.4 Estimating Potential Effects in the Receiving Environment

Sublethal toxicity test results can also be used to predict the likelihood of zones of potential effect in the receiving water. It is desirable to make only one such prediction from a series of sublethal laboratory tests. Three options for selecting a concentration to represent the test results, for comparison with the potential zone of effect, are recommended.

- Use the lowest IC25 available from sublethal toxicity tests conducted on the effluent of a given facility.
- Use the geometric mean of the sublethal IC25s for the effluent of a given facility.
- Use the *Predicted Minimum Toxic Concentration (PMTC)*, as calculated by a specified statistical procedure.

Lowest IC25. The first option is simply a comparison of (a) the lowest IC25 or EC25 from a series of tests, with (b) the concentrations expected in the plume where the effluent mixes into the receiving water. One benefit of this approach is that an individual test endpoint can be assessed in relation to the series of sequential results from the same effluent; temporal variation in effluent quality can be established. However, caution must be exercised to avoid using the lowest endpoint if it is a non-representative outlier. That would lead to a poor, overly conservative estimate of the potential zone of effect. It is essential to confirm the validity of the test if this approach is chosen. All the requirements for a valid test should be at least be met, and preferably exceeded, and the endpoint should show a narrow 95% confidence interval. This option is recommended if the facility has ≥ 10 sets of results for a battery of sublethal toxicity tests, and the tests reflect the current processes and effluent treatment.

An insufficient number of toxicity tests was one problem during the first cycle of pulp and paper EEM, which required only four sampling periods. That was a limitation in comparing IC25s with concentrations in the plume. Four data-points do not give a good representation of the variability in quality of mill effluent, and accordingly, picking the lowest IC25 was not a satisfactory procedure.

For most pulp and paper mills, there will be six sets of sublethal toxicity tests for each of three species, during Cycle Two of EEM. The previous set of four sublethal toxicity tests from Cycle One might be added to the series if effluent quality had not changed substantially. The total number of tests could then rise to 10, which is still a fairly small number considering the usual variation in results. If a larger number of samples had been tested, some results would be expected to show less toxicity than the least toxic of the 10 tests, while other additional samples would be expected to show greater toxicity than the most toxic of the ten. In many cases, only the six sets of results from Cycle Two will be useful to reflect the current level of effluent treatment, because secondary treatment was installed during the latter half of the first cycle.

To partially compensate for a low number of sublethal test results (i.e., <10), one of two slightly more sophisticated approaches can be used. One method is to use the geometric mean of the IC25s generated during a given period of testing. The second method involves a statistical procedure which uses the available endpoints to calculate the lowest IC25 likely to occur with repeated testing of that effluent under similar conditions. Both of these mathematical methods for characterizing the sublethal test results require at least three IC25s for each discharge location (or three EC25s, for the salmonid embryo test). These methods are described in the following text.

Geometric mean. This statistic can be used to characterize a set of test endpoints for a given species at a discharge location. The geometric mean is calculated using all the IC25s for the species (or EC25s, for salmonid tests) [47]. The standard deviation is also calculated using the logarithms of the endpoints. The anti-logarithm of the result should be taken, and that arithmetic value for the standard deviation should be reported. This method can be used for three or more endpoints in a series.

The PMTC. In situations where there are fewer than ten IC25s (or EC25s) in a series, a statistical procedure can be used with those available endpoints, to predict the lowest IC25 that would likely be encountered in a much larger series of tests. This estimated value is called the *Predicted Minimum Toxic Concentration (PMTC)*, and is defined as the concentration of effluent, below which only 5% of samples would be expected to have a deleterious sublethal effect, as estimated at the 95% level of confidence. As an alternative description, the PMTC is an estimate of a low value of IC25, such that at least 95% of the population of IC25s would be higher (less toxic), with a 95% confidence level. The procedure follows standard probability relationships on the assumption that the population has a lognormal distribution. The procedure has been modified from USEPA (1991a), and is described in the following text. It can be used for three or more endpoints in a series.

---

[47] The geometric mean can be calculated as the nth root of n numbers multiplied together. For the second cycle of EEM, it will usually be the 6th root of six IC25s multiplied together. Alternatively, the logarithms of the six IC25s can be added together, the sum divided by six, and the antilog of the result is the geometric mean. The logarithm of each IC25 is also used in calculating the standard deviation. Alternatively, natural logarithms could be used throughout.

To calculate the PMTC, the geometric mean and the logarithmic standard deviation are calculated, as in the preceding technique. A *compensation factor* is then obtained from Table D2 in Appendix D, for the 95% level of confidence. The mean, standard deviation, and compensation factor are used to calculate the PMTC, by the method given in the following example. This statistical approach is fairly conservative in that the estimate is affected by variation in effluent quality and limited numbers of test results. If even greater conservatism were desired, a compensation factor for 99% confidence could be taken from Table D2 and used in calculations.

A worked example of calculations for the *Selenastrum* test is shown in Appendix D. In this example, testing of a mill effluent on four occasions resulted in measured IC25s of 15, 17, 25, and 30% (Table D1), and the resulting estimate of PMTC was 3.7%. The appropriate compensation factor from Table D2 is based upon the number of measurements of IC25; in this case, the number is four so the compensation factor is 5.369. The compensation factor is then used to multiply the logarithmic standard deviation, and the result is subtracted from the logarithm of the geometric mean. The antilog is taken, of the number resulting from that subtraction, and that gives the estimate of the PMTC. As seen in the example, the PMTC for *Selenastrum* is lower (i.e., more toxic) than the lowest measured IC25 since the statistic considers the degree of variability of sublethal toxicity of the effluent, and it also factors in the low number of data-points. To repeat, this PMTC is an estimate of the greatest sublethal toxicity to *Selenastrum*, that would be expected to occur or be exceeded (i.e., even more toxic), 5% of the time, in samples of effluent from this source.

The PMTC was calculated for most Canadian pulp and paper mills as part of the review of data from Cycle One. Mills with only primary treatment showed wide variation in toxicity of effluent, and consequently showed low values for PMTC. The small number of toxicity tests (only three or four data-points) also resulted in lowered values for PMTC. For some mills with primary treatment, the difference between the mean IC25 and the PMTC was six orders of magnitude. For mills with secondary treatment, the difference between mean IC25 and PMTC was typically only three orders of magnitude. Better effluent treatment was the main reason for the reduced difference between mean IC25 and PMTC, although an increased number of data-points also had a considerable influence. During the review of data from Cycle One of the pulp and paper EEM, the very low values for some PMTCs were condemned as unrealistic by some groups concerned with environmental effects. In fact, the estimates were statistically realistic and pointed to potentially unsatisfactory situations in which there was clearly a need for larger numbers of toxicity tests, and/or less variability in measured toxicity, in order to validate the statistical estimates.

The Canadian findings can be compared with results of more extensive testing at nine pulp and paper mills in the United States all of them with secondary treatment. The data on sublethal toxicity to *Ceriodaphnia* were made available to the Canadian industry/government Expert Working Group on Laboratory Toxicology during the review of data from Cycle One. Each U.S. mill had from nine to fifteen IC25s. The mean IC25 and the PMTC for each U.S. mill were calculated, and usually differed by one order of magnitude or less. This close concordance was likely because: (1) each of the U.S. mills had relatively stable effluent quality; (2) the *Ceriodaphnia* tests were conducted by a laboratory specializing in the test; and (3) a relatively large number of tests were available for each mill. Some Canadian mills should have ten IC25s for each sublethal toxicity test by the end of Cycle Two, at which time a similar trend might be seen for the PMTC to rise, and more closely approach the geometric mean IC25.

The Expert Working Group on Laboratory Toxicology raised two areas of concern about using the PMTC to estimate a potential zone of effect, as part of EEM for the pulp and paper industry.

1. *Normality of data and small sample size*. An assumption in the compensation factor used for estimating the PMTC is that the data should have a distribution that is approximately normal. An attempt was made to analyze the *Ceriodaphnia* data from U.S. pulp mills, but the sets of results did not contain enough values to determine whether normal distributions existed. A statistician advising the working group felt that normality of the data was unlikely to be a problem in this situation, and that the cause of the low values for PMTC was the small sample size (a maximum of four test results in Cycle One).

2. *Source of variation*. Two main sources of variation can influence the PMTC. These are (a) variation in the response of the test system, and (b) change in effluent quality because of changes of process within the mill, in the treatment system, or seasonal influences. Test system variability can result from response of the organism or variation in technique. In Cycle One of the EEM for the pulp and paper industry, it was apparent that a consistent standard of quality was not in place for the sublethal toxicity tests, and too much flexibility was allowed in the level of quality assurance to be used by the consulting laboratories that did the tests. For Cycle Two, the level of internal quality control has been defined through amendments in the official test methods, and changes in the requirements for EEM. A strong recommendation has also been made by Environment Canada, that mills should contract only those laboratories having accreditation in sublethal testing from an external program (e.g., CAEAL or MEQ). Those changes should assist in controlling within-laboratory variation. The Expert Working Group also recommended that IC25s generated after a major change in mill processing or waste treatment should be considered separately, and not combined with IC25s determined before the change, for purposes of estimating the PMTC.

The Expert Working Group also sought advice on seasonal variation in effluent quality at Canadian pulp and paper mills from effluent treatment engineers, who provided a consensus that there is little change in quality between summer and winter, as long as the treatment system is well managed (EC, 1997c).

Comparing the laboratory results with a plume. Under the EEM program for pulp and paper, each mill must map the plume of effluent discharge into the receiving water, and must show where there is a decrease to 1% concentration. The plumes are mapped by field measurements and/or modelling, for critical periods of low flow in rivers, or poor mixing conditions in lakes, estuaries, and oceans. The maps of effluent plumes make it a relatively straightforward exercise to compare the concentrations expected in the receiving water with the sublethal toxicity measured for the effluent.

The toxicity criterion used in the comparison would be obtained by one of the three methods described previously, i.e., the lowest measured IC25, the geometric mean IC25, or the PMTC. Whichever concentration is used, it indicates the border of a region in which a specified effect would be found for a particular species. Therefore, a "potential effects zone" can be defined as that area of the plume in which concentrations exceed the selected criterion of sublethal effect. The potential effects zone for each species can be mapped onto a diagram of the effluent plume, to illustrate whether it is larger or smaller than the area of plume defined by the isopleth for 1% concentration of effluent (see Appendix D).

5.2.5 Understanding Effects in Multiple Discharge Situations

Estimates of sublethal toxicity can help in understanding the relative contribution of diverse industrial/municipal discharges to effects on aquatic organisms in the receiving water, whether the discharges are from point or non-point sources (e.g., landfill leachate, bark pile runoff, acid rock drainage). Each contribution to an the observed environmental effect can be estimated, given data on sublethal toxicity, discharge flow, and features of dispersion into the receiving environment. If plumes from different discharges overlap, more effort is required to distinguish the toxic contributions of the discharge sources. Samples of surface water from key locations in the near-field exposure areas could be tested, to estimate the combined toxic contribution of the sources.

The USEPA (1991a) recommends the following procedure for separating individual toxic contributions, in multiple-discharge situations.

1. Conduct a battery of sublethal toxicity tests on samples collected from each discharge source. Use standard laboratory for test dilutions and control, or uncontaminated site water. This estimates the absolute sublethal toxicity of each discharge. Repeat the sampling and testing on any discharge known to be variable in toxicity, to obtain an estimate of the degree of variability.

2. Conduct a parallel battery of sublethal toxicity tests for each discharge to a river, using water collected directly upstream from the point of discharge for dilution and control. For lakes or estuaries, carry out the parallel battery of tests by collecting control/dilution water from outside the zone immediately affected by the discharges. Separate and simultaneous controls must be run using standard uncontaminated water as a quality assurance measure.

   It must be recognized that "upstream" sources of control/dilution water might already be contaminated by other effluent discharges or sources of toxicants. Accordingly, the upstream dilution water might contribute to significant effects on growth or reproduction, in concentrations of the effluents being studied, or even in control vessels. This would not invalidate the results since the purpose of the investigation is evaluate the relative contributions of discharges to the total toxicity of the receiving water.

As an additional step, confirmation of the relative contribution of discharges can be achieved by conducting sublethal toxicity tests on samples of water from the waterbody receiving the discharges (so-called "ambient" tests). This can aid in:

a. confirming whether or not an effluent has a measurable toxicity after mixing into the receiving water;

b. estimating the persistence in the receiving water, of toxicity from all contributing sources; and

c. determining the combined toxicity resulting from the mixing of all point- and non-point sources, as an estimate of the overall effect on the receiving environment.

Testing samples of surface water which receives discharges or toxicants from multiple sources should be done synoptically and ideally during low-flow or worst-case periods. At a minimum, sampling should be carried out over as short a period of time as possible (e.g., 1 or 2 days). Repeated rounds of sampling and testing would be desirable if the toxicity of the discharges were variable. The above guidance on multiple-source toxicity assessment is based on the eight site investigations conducted under the USEPA Complex Effluent Toxicity Testing Program (Mount and Norberg-King, 1985, 1986; Mount *et al.*, 1984; 1985; 1986a, b, c; Norberg-King and Mount, 1986). The results of that program provided valid findings, relevant to the Canadian metal mining industry and the Aquatic Effects Technology Evaluation (AETE) Program, according to a review of the studies and subsequent re-analyses and independent critiques (Sprague, 1997).

Another method of comparing inputs of sublethal toxicity from different effluents or other sources is to calculate the *Toxic Emission Rate (TER)* or the *Toxic Emission Factor (TEF)*. Both assess relative *loading*, i.e., the volume of material discharged, as well as its degree of toxicity. Using the formula TER = (Sublethal Toxic Units) × (discharge flow), the toxic loadings can be rationally compared for discharges of different volumes and different toxic strengths. A further refinement considers the toxic loading per unit of industrial production, using the formula TEF = (Sublethal Toxic Units) × (discharge flow)/(facility production rate). These concepts are detailed in Section 4.3.3. Dispersion of effluent in the receiving water must also be considered, to fully use these concepts of toxic loading.

5.2.6 Integrating Toxicity Results into Environmental Monitoring

Any program of EEM must consider the total spectrum of information from biological surveys and chemical measurements along with the extrapolated results for sublethal toxicity of effluent(s) or receiving water(s). Any single tool might yield divergent information.

Certain other points should be considered when sublethal toxicity data are used to delineate a "potential effects zone".

1. *Background toxicity*. Some of the previous descriptions assume an absence of significant upstream contributions of toxicity. That assumption would be erroneous if there were overlapping plumes, and the potential combined effects would have to be accommodated in any interpretation. If there were additional confounding discharges and possible overlapping plumes, the guidance in Section 5.2.5 should be followed.

2. *Type of receiving water*. The pH, hardness, DOC and other characteristics of the receiving water could increase or decrease the toxicity of the effluent compared to tests with laboratory water. For metal mining studies, one option for reducing such modifying influences is to collect uncontaminated water from upstream or from a reference site, and use it for control and dilution water as addressed in the technical guidance document for EEM in the mining sector. That approach was investigated as part of the Aquatic Effects Technical Evaluation Program

(AETE, 1997; 1998b). It was concluded that there was limited value in using uncontaminated site water, especially considering the additional cost of collecting it, often from locations inaccessible by road. In EEM for pulp and paper mills, use of local site water in sublethal toxicity testing is not recommended by the Expert Working Group, since the quality of the receiving water would not be a major influence on the action of organic toxicants, typically found in pulp and paper effluents (EC, 1997c).

3. *Plume uncertainties*. Calculations of dilution might be difficult or inaccurate, or the position of the mixing zone might be variable. In such cases, resident biota would be associated with uncertain or variable levels of exposure.

4. *Species differences*. The potency estimated in the laboratory might not be seen in the receiving water, if the test organism is more sensitive than species in the local environments.

The comparison of toxicity tests with results of chemical and biological surveys should take into account the nature, location, and magnitude of changes in the receiving environment. At a minimum, observed effects should correspond spatially with the direction of flow of the effluent plume and should diminish in magnitude with distance from the outfall [48]. Dilution in the waterbody must be explicitly rationalized. Biological effects attributed to the effluent at a given site should correspond at least approximately with the estimated strength of the effluent. The biological effect assumed to result from exposure to effluent should not extend very far beyond the points where the effluent plume would decrease below the lowest IC25. Sublethal toxicity predictions often seem to correlate well with observations on sedentary organisms such as the benthic invertebrates [49]. Substantiation that a fairly good relationship exists with survey results for benthic invertebrate communities is evident in data from Cycle-One EEM programs at 17 pulp and paper mills in the Atlantic Region. In 12 out of 12 marine studies, sublethal toxicity test data correctly predicted that there would be effects on the benthic community. In one out of four freshwater studies, the toxicity tests incorrectly predicted benthic effects which were not found. The relationship was not as strong, between sublethal toxicity and health indicators for adult fish caught in surveys, mainly because of inadequate catches and inability to confirm exposure to mill effluent (Parker and Smith, 1997).

5.2.7 Mixing Zones

Mapping the area of potential effects begs the question of what is an "allowable" zone of sublethal effects. No national guidance exists on allowable mixing zones in Canada, since establishment of mixing zones is currently not possible under the federal Fisheries Act. Each province, however, can have its own policies or definitions. Most provincial guidance does not differentiate acute from chronic effects, whereas EEM deals specifically with sublethal (≈ chronic) effects.

British Columbia defined a mixing zone for pulp mills, describing it as an *initial dilution zone* in which water quality guidelines might be exceeded (Pollution Control Board, 1977), i.e., the zone allows sublethal effects to prevail in a limited area. The B.C. Pollution Control Board functioned under the auspices of the provincial Ministry of the Environment. It defined the initial dilution zone as extending up to 100 metres horizontally, but not exceeding 50% of the width of the waterbody. For lakes, estuaries, marine waters, and multiport discharges, the distance of 100 m applied in all directions, while for rivers and streams, it applied downstream of the discharge point.

Quebec has had size limits on mixing zones since 1980, and requires absence of sublethal toxicity at the edge of the zone. Setting such toxicity limits is a prerogative of Canadian provinces. Taken together, the following size limits applied by Quebec would appear to be protective for smaller waterbodies, and would also specifically guard against immense pollutant loading to very large rivers with great dilution, such as the St. Lawrence. The limits might be considered when interpreting results from EEM.

---

[48] There could be exceptions to such a generalization. For example, solids containing toxicants might be carried downstream and then deposited, and might then create an area of local increased effect.

[49] A review of 63 studies, suitable for comparing results of sublethal effluent tests and effects found in the aquatic communities receiving the effluents found 53 cases of agreement and 10 cases of disagreement, for an overall 84% rate of agreement (Sprague, 1997). That degree of correlation might not hold for pulp mills if historical discharges had resulted in significant zones of deposited wood fibre and bark, because habitat alteration might be the dominant factor in determining the zone of potential effects.

- <300 m in length
- width less than half the width of the river, to a maximum of 50 m
- maximum dilution of 1:100 (i.e., the edges of the zone are defined by a decrease to 1% concentration)

Alberta has established a two-stage zone of mixing which is based on *in-stream guidelines,* chemical concentrations listed in a procedures manual of Alberta Environmental Protection (AEP, 1995). The chemical guidelines are drawn from values published by Alberta (AEP, 1978) and in the "Gold Book" of the USEPA (1986). *Acute guidelines* are intended to protect aquatic organisms from short-term lethal action based on a one-day average concentration, and the so-called *chronic guidelines* are intended to protect from sublethal effects, based on average concentration during four days. There is a comprehensive list of flow-related and general restrictions on the size of mixing zone that must be met, to protect other uses of the water and the waterbody as a whole. (For example, the mixing zone must not kill or create a barrier for passing pelagic organisms, must not impinge upon drinking water intakes, and must not create problems with bioaccumulative substances). In addition, the following absolute maxima for sizes of mixing zones are specified.

- The zone in which the acute guidelines might be exceeded must not exceed 30 m in any direction.

- The surrounding plume, in which chronic guidelines might be exceeded, is limited to one-half of the width of a stream, or a length that is 10 times the width of the stream, whichever is more restrictive at low flow.

The following two summaries describe the guidelines for "sublethal mixing zones" published by the USEPA Region VIII (1994).

- Flowing water (river or stream): sublethal mixing zone limited to a width of one-half of the cross-sectional area of the river, and a length that is 10 times the width of the river, whichever is more limiting at low flow.

- Standing water (lake or reservoir): sublethal mixing zone limited to 60-m radius or 5% of lake surface area, whichever is smallest.

The definitions from USEPA Region VIII would require some additional wording before being used in Canada. For example, to avoid vast mixing zones in large rivers, an absolute limit of length should be added, such as the Quebec maximum of 300 m. Coastal marine waters might be included in the second U.S. item, with a limit of 60 m for the radius.

5.3 Assessing Sediment Toxicity for Dredging and Disposal at Sea

Key Guidance

- *Disposal of non-hazardous materials at sea is regulated under the Canadian Environmental Protection Act (CEPA).*

- *In the Canadian procedure, materials for disposal are first assessed by chemical screening, and then if appropriate, by a battery of toxicity tests.*

- *Monitoring uses a tiered approach with similar testing, under a formal set of procedures.*

- *Toxicity tests should include a reference sediment. Its acceptability is judged mainly by similar grain sizes, total organic carbon, and lack of toxicity. Proper selection is difficult, and procedures are suggested to aid selection and subsequent interpretation of results.*

- *Relative sensitivity of toxicity tests for sediment appears to be: bacterial luminescence (solid-phase or extraction) ≥ echinoid fertilization > amphipod mortality > bacterial luminescence (porewater) = polychaete growth and survival. Acute lethality to an infaunal amphipod appears to be a sensitive test that generally provides a good indication of conditions in the actual environment, although species differ in sensitivity.*

5.3.1 Canadian Approach

Part VI of the Canadian Environmental Protection Act (CEPA) regulates disposal at sea by means of permits. Disposal at sea is allowed only when it is the environmentally preferred and practical alternative for inert, non-hazardous, uncontaminated materials (CEPA, 1988). Dredged material from channels and harbours is the most common waste. Other materials include fish waste and animal feed which cannot be used for fertilizer, scrap metal, and old vessels. Disposal sites are monitored for compliance with permit conditions, and to verify that the process of review and site selection was suitable for environmental protection.

Canada's approach is based on the London Convention 1972, an international agreement among more than 70 countries to control pollution from dumping of wastes into the seas (IMO, 1990; also summarized in EC, 1993b; 1995b). Canada evaluates materials for disposal at sea by using the Convention's *Waste Assessment Framework*, which advocates a precautionary approach, full consideration of options, and use of scientific procedures to characterize the waste and to decide on its suitability for disposal (Figure 9; EC, 1995b). If defined chemical screening limits are exceeded, toxicity tests are used to characterize waste as part of application for a permit. Permit decisions are followed by monitoring to test the impact hypotheses derived during review (EC, 1993c), and toxicity tests might also be used as part of this monitoring. Results of monitoring feed back into site selection for possible revision of predictions, granting of permits, and their conditions (Figure 9).

For **characterizing and assessing** materials for disposal at sea, the Canadian tiered approach uses chemical analysis and sediment toxicity tests (EC, 1993b). Chemical variables include PAHs, PCBs, metals, and dioxins, and wastes with a specified substance above the screening level would require toxicity testing. The battery of five biological tests would include the bacterial luminescence test, a 10-d amphipod survival test, a test of fertilization success with sea urchins or sand dollars, (EC, 1992c,e,f), a survival and growth test with polychaete worms (EC, in prep., b), and a bioaccumulation test (USEPA,1993). These would all be pass/fail tests, without dilution of the waste material (EC, 1993b).

The Canadian **monitoring** approach is outlined in the Interim Monitoring Guidelines for Ocean Disposal (EC, 1993c), in the related Biological Monitoring Guidelines (EC, 1994b), and in interim guidelines for interpreting data (EC, 1996b,c). This is a tiered approach, with parallel efforts on (a) sediment chemistry, (b) sediment toxicity tests, and (c) benthic community surveys (Figure 10). Further *in situ* biological investigations could be undertaken in relation to site-specific impact hypotheses, or could be triggered by previous assessments of the sediment.

In both the assessment and monitoring of sediments at disposal sites, it is a conundrum that most methods for collecting samples cause some disturbance of the vertical gradients in the substrate (EC, 1994c). Mixing or sieving of samples in preparation for toxicity tests completely destroys the gradients. Such disturbance can alter physicochemical characteristics that determine bioavailability of toxicants. A most important vertical gradient is the redox state, often very steep and driven by decomposition which uses oxygen faster than it can diffuse downwards (Burton, 1991). Other gradients in chemical composition parallel redox (see Section 4.15.1). Sediments also develop physical gradients as deposition of particles is followed by re-sorting from differential settling, currents, and animal activity.

Accordingly, the toxicity measured in a laboratory test is more comparable to sediments immediately following their disposal. As gradients develop over time in discarded sediments, they can become more or less toxic. This partiality of the approach must be kept in mind when evaluating toxicity data.

5.3.2 Reference Sediments

This section focuses on testing sediments rather than other materials, because dredged sediments are the most common material disposed of at sea. Results from a toxicity test on sediment should be compared for significant difference from either a control sediment, a reference sediment, or both [50]. Section 4.4 discusses the interpretation of data from sediment toxicity tests, and Ingersoll (1995) provides a review.

---

[50] Whether the results from a test are compared with control or reference should depend partly on the purpose and design of a given program. Sometimes both are desirable, for different reasons. Control sediment can be obtained from the site where organisms were collected or it can be sediment used in culture. Comparison with a control is useful, since it is known to be uncontaminated and conducive to the health of test organisms.

Reference sediments are usually collected from a site near that of the test sediments, but outside the influence of pollutants, so they are presumed to be clean. A reference sediment must be used for comparison if a field-collected sediment is being tested with marine or estuarine amphipods (EC, 1992e). The rationale is that a reference sediment is more likely to be similar to the test sediment in variables such as grain size. It is desirable to eliminate or minimize such variables as potential causes of biological effects, in order not to confound the effect of toxic substance(s).

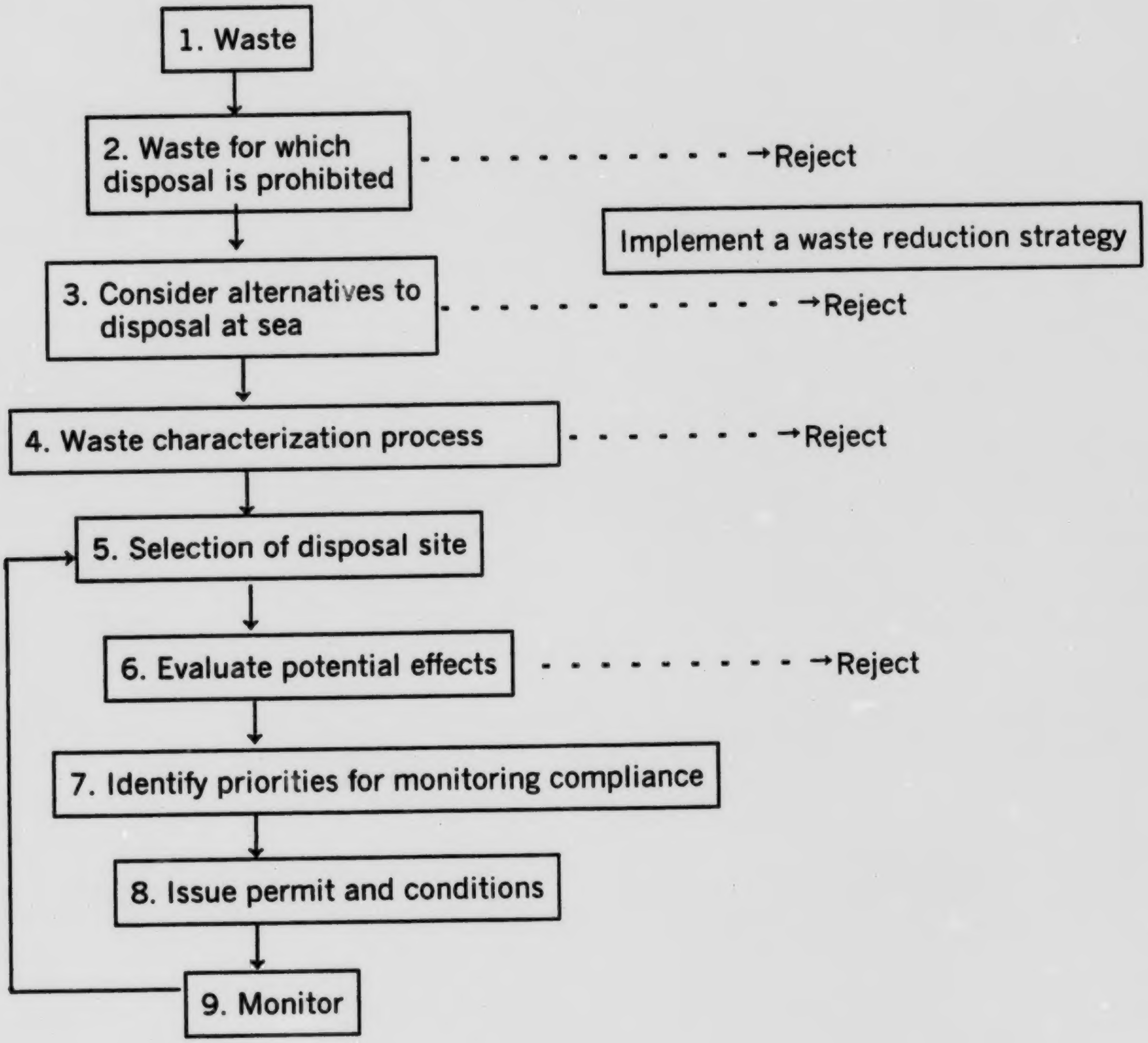
1. Waste
2. Waste for which disposal is prohibited
Reject
Implement a waste reduction strategy
3. Consider alternatives to disposal at sea
Reject
4. Waste characterization process
Reject
5. Selection of disposal site
6. Evaluate potential effects
Reject
7. Identify priorities for monitoring compliance
8. Issue permit and conditions
9. Monitor

Figure 9 **A Framework for Assessing Wastes for Disposal at Sea.** A simplified schematic of the approach recommended by the London Convention 1972 (IMO, 1990). From Chevrier and Porebski (1993).

Choosing a site for collecting reference sediments is covered in Section 4.11.1. Sometimes there are difficulties in selecting a location that is near the site of the test sediment, has a sediment of similar characteristics, but is uncontaminated. In a given locale, contaminated sediment with fine grain sizes will tend to settle in quiescent areas; nearby clean areas can have faster currents (hence the lack of settling of fine contaminated sediment) and correspondingly coarser grain sizes.

As another very pragmatic difficulty, a field sampling program (and perhaps the toxicity testing) might be completed before any physicochemical analyses were available, to aid in choosing reference sediments. Samplers should attempt to match sediments by noting apparent characteristics in the field and rejecting obviously unsuitable samples. Samplers might also attempt to obtain a series of reference samples, with particle-size distributions bracketing those of the test sediment(s). One of the reference sediments might be a good match; if all were tested, the biological effects attributable to grain size should be apparent as differences among the reference sediments.

Similar sediments might not be available nearby, or they might all be contaminated. In that case, sediments from a nearby similar coastal zone could be acceptable if (1) the location had comparable conditions (e.g., another cove farther along the coast), and (2) the samples had similar grain sizes, total organic carbon (TOC), and lacked innate toxicity. Particular attention should be given to the proportion of fine particles (<63 µm) because those sizes bind most contaminants and are also important for determining suitable habitat for benthic organisms. Unfavourable distributions of particle size can also create stress in test organisms, with a measurable response. The bioavailability of nonpolar organic chemicals is governed mostly by total organic carbon, rather than particle size (see Section 4.11.3).

Reference sediments can be toxic because of ammonia, which is the most common natural toxicant in sediments, especially for waterbodies that are eutrophic or have high TOC (Sarda and Burton, 1995). Sulphide is also a common natural toxicant in sediments (Knezovich *et al.*,1996). Other characteristics such as salinity and redox potential are of less frequent importance, but can be important at specific sites such as Pacific coast fjords (see Sections 4.11.1 and 4.15.1). Potential reference sediments might also contain toxic substances of which the investigators were unaware. This can create a dilemma because toxicity in the test sediments might be ascribed to the known anthropogenic pollutants, or to a natural toxicant shared with the reference sediment.

If a reference sediment proves toxic, the current Canadian practice in testing for disposal at sea is to use results from the control sediment. In tests with amphipods, results for the reference sediment are not used if mortality is 20% greater than in the control test and is statistically different. (Criteria for designating a test sediment as toxic are given in Section 5.3.3).

It is desirable to obtain a satisfactory reference sediment for testing, and the most convenient way to do that is to sample a number of plausible locations (as suggested previously). That approach should certainly be followed if it is anticipated that there might be difficulties with the reference sediment. It would be a good precautionary approach in any survey program, in case there were unexpected difficulties. It would usually be more economical to collect from additional sites at the time of the survey, rather than go back later to find a suitable reference sediment. The approach would involve the following two steps.

1. Collect several potential reference sediments from different sites thought to be suitable. As suggested previously, inspect the general characteristics of each sample as it is collected, for similarities to the test sediments (particle size, organic content, and any other features easily observed). Continue until it appears that several suitable reference sediments have been collected.

2. In the laboratory, pre-screen the potential reference sediments for particle size, organic carbon, sulphides, and toxicity by a rapid bacterial luminescence test. Select the sediment, or sediments, that seem most suitable.

Occasionally, for a major program on an important disposal site, toxicity of reference sediments might warrant more extensive investigation. There might be justification for identifying the toxicant(s) by applying some of the techniques of toxicity identification (TIE, Section 4.7). For example, the toxic contribution of ammonia in a reference sediment could be investigated or corroborated by adding zeolite to detoxify it (Besser *et al.*, 1998). Such detailed investigation would be

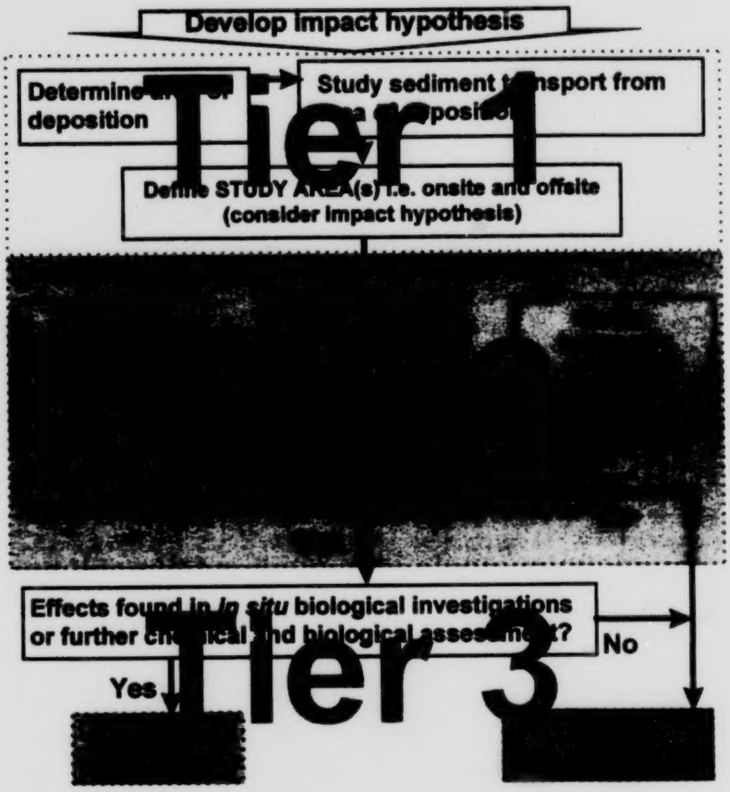
Develop impact hypothesis
Determine ... deposition
Study sediment transport from ... deposition
Tier 1
Define STUDY AREA(s) i.e. onsite and offsite (consider impact hypothesis)
Effects found in in situ biological investigations or further chemical and biological assessment?
No
Yes
Tier 3

Figure 10 **Canadian Approach for Monitoring Marine Sites used for Disposal of Wastes.** Three successive tiers of activity are indicated. From Chevrier and Topping (1998).

time-consuming and costly, however, and it might be more profitable to focus efforts on other parts of the program or on finding a more suitable reference sediment.

5.3.3 Assessing Waste Sediments for Toxicity

In current interpretation of the Canadian Environmental Protection Act (CEPA), a battery of toxicity tests can be used to demonstrate that a material is "rapidly rendered harmless" for disposal at sea (EC, 1994f). The tests with sediment use amphipods (EC, 1992e), echinoids (EC, 1992f), photoluminescent bacteria (EC, 1992c), and a bioaccumulation test with the mollusc, *Macoma* sp. (USEPA, 1993). If the material proposed for disposal contains substances in excess of regulated or guideline levels, the results of toxicity tests have a definitive role (EC, 1994f). If the material passes all four biological tests it is considered acceptable for open-water disposal. If it is not acutely lethal to amphipods, but fails one of the sublethal tests or the bioaccumulation test, the material may be acceptable for disposal with capping or confined aquatic disposal. If the material fails the acute lethality test with amphipods or two or more of the other tests, disposal at sea is not permitted.

Interim pass-fail criteria have been defined for the toxicity tests used in Canada for deciding suitability for disposal at sea (EC, 1996b). In tests with amphipods, the criteria differ slightly depending on the species used. For *Amphiporeia virginiana*, the sediment is considered toxic if it has mortality greater than 50%, and the reference sediment has mortality ≤30%. For *Eohaustorius washingtonianus*, the comparable criteria are >45% and ≤25%. For *E. estuarius* and *Rhepoxynius abronius*, the criteria are parallel but slightly different at >40% and ≤20%. For bacterial luminescence (Microtox® solid-phase test), the material is considered toxic if a concentration of

<1000 mg/kg results in a 50% inhibition of light production after five minutes of exposure. For the echinoid test, the material is considered toxic if it is associated with a statistically significant decrease in fertilization of at least 25% compared to the control. The echinoid test compares porewater with a control water. For bioaccumulation in the mollusc, the criterion is "a statistically significant difference in the tissue concentrations" of organisms exposed to the test sediment, compared to those exposed to reference sediment (EC, 1994f, 1996b; Chevrier and Topping, 1998).

The purpose of using a battery of toxicity tests on waste, rather than one test, is to strengthen the evidence before a material is designated as suitable for disposal at sea. A given sample of waste is unlikely to be equally toxic in all the toxicity tests. There need be no contradiction in such results, because different species differ in their sensitivity and might not even be responding to the same toxicants. Indeed, unequal sensitivity among species is the reason for using a battery of tests, as emphasized throughout this document (see Section 2.3).

The following examples illustrate some of the differences in sensitivity that can be seen among the species used in sediment tests. It should not be forgotten that sensitivity is only one of three important aspects of a toxicity test. A second aspect is precision of the test, and Section 3 describes quality control which can be related to precision. Another aspect is accuracy which cannot be directly assessed in toxicity tests, but can be judged by comparison with findings in more complex biological systems, and such validation is described in Section 4.1.

An example of variation in results obtained from different sediment toxicity tests can be taken from four tests carried out at each of seven sites in Halifax harbour (Tay *et al.*, 1992). Bacterial luminescence tests were the most sensitive of the four methods, when used with solid phases and solvent extracts. This method detected toxicity at virtually every site including a relatively clean reference site, which suggested a response to natural factors. The porewater version of the bacterial test indicated fewer toxic sites, but was a good estimator of toxicity for water-soluble substances.

The next most sensitive test was percentage survival and reburial of the Pacific amphipod *Rhepoxynius abronius*, a species which burrows in sediments. Tests with the Atlantic amphipod *Corophium volutator* showed responses to only a few of the samples, presumably because it is a tube-builder which has less direct contact with sediment. The test based on survival and biomass of juvenile polychaete worms (*Neanthes* sp.) was the least sensitive and did not indicate toxicity at most sites.

Similar findings are reported for eight sediments from a disposal site in Saint John Harbour, New Brunswick (Tay *et al.*, 1993). The solid-phase bacterial luminescence test was again the most sensitive test, while the bacterial test with porewater and the polychaete test gave few positive results. Whole-sediment tests with the infaunal amphipod *Amphiporeia virginiana* and *R. abronius* were similar to each other in results, but a fertilization test with a sea urchin *(Lytechinus pictus)* was more sensitive than either amphipod test. Sea urchin fertilization was the most sensitive of five tests used on sediments from Tampa Bay (Carr *et al.*, 1996).

Considering the foregoing results, the order of sensitivity appears to be as follows, among toxicity tests proposed for testing sediments for disposal at sea: bacterial luminescence (solid-phase or solvent extraction) ≥ echinoid fertilization > amphipod mortality > bacterial luminescence (porewater) = polychaete growth and survival (Tay *et al.*, 1992; 1993). Among the solid-phase tests, the acute lethality tests with certain species of infaunal amphipod, notably *R. abronius*, have been widely tested and appear to be sensitive, reasonably reliable, and applicable to most coastal sites. Choice of the species of amphipod is important, however, because there are significant between-species differences (Thursby and Schlekat, 1993). Variability can occasionally be a problem in all kinds of tests, but results for amphipods tend to be consistent, providing reasonable statistical power for applying the results (Anderson *et al.*, 1998).

Despite these generalizations on the usefulness of amphipod tests, any program to establish suitability for disposal at sea must follow Canadian practice which uses five sediment tests as described previously (EC, 1994f). Any type of investigation, official or not, should continue to follow the principle of using a battery of diverse tests. For example, only 50% of the test results with *R. abronius* in Halifax harbour concurred with the solid-phase bacterial luminescence test. It would be imprudent, therefore, to rely entirely on either of the tests without confirmation from another. Amphipod tests might show a bias because they were influenced by particle size. For example, tests with *R. abronius* in California were highly sensitive to the samples and were relatively high in precision, but results were most strongly correlated with particle size and other variables in the sediment (Long *et al.*, 1990). Similarly, several hundred tests with *R. abronius* on uncontaminated sediments of the Alaskan coast indicated that grain size played an important role in survival (Pillard *et al.*, 1993). The solid-phase bacterial luminescence test has also been found to be sensitive to the particle sizes of sediments being tested (Benton *et al.*, 1995). Tolerance for particle size is known or partly understood for some marine organisms (especially amphipods); guidance might be found in the documents describing test methods, or it might be found in the literature (DeWitt *et al.*, 1988; 1989).

There are also notable differences between species of amphipods. In the California study by Long *et al.* (1990), the amphipod *Ampelisca abdita* appeared relatively insensitive to the sediment samples, compared to *R. abronius*, but showed results that were highly correlated with concentrations of toxic chemicals, and with no other variables.

Procedures for checking the quality of the toxicity tests are outlined in all Environment Canada test methods, and in Section 3 herein.

5.3.4 Monitoring Disposal Sites for Toxicity

Monitoring disposal sites is an integral part of controlling any marine disposal under the Waste Assessment Framework. Monitoring is intended to verify that the assumptions and predictions made during the review of a disposal application, were correct and sufficient to protect the environment (EC, 1993c;1995b). Issuing a permit is tantamount to stating an hypothesis that the material will have no significant detrimental effects on the environment. The primary hypothesis is that toxicity of the disposed sediment will not significantly alter the biological community near the disposal site. Decisions on conditions at a disposal site will use interpretations of chemical and toxicity data, based on the criteria and guidelines from the assessment phase for issuing permits. The interpretation should also be based as much as possible on comparisons with control data, both spatial and temporal. The decisions should also take other factors into account, such as the likelihood of dispersal beyond the boundaries of the site (EC, 1996c).

The information presented in Sections 5.3.2 and 5.3.3 should also be considered when interpreting monitoring tests. Interpreting the impact hypotheses can be assisted by using the same species as were used for waste assessment (Section 5.3.1). Use of a battery of tests is essential, and current Canadian practice is to use at least three tests, including those used during permit review. A species of local relevance might be selected if a choice is provided in the standard test method. A reference site should be considered as an aid, in comparing biological conditions at the disposal site.

5.4 Environmental Assessments

Key Guidance

- *Environmental assessment (EA) is a scientific prediction of the effects of a particular project or development on the natural environment. It includes public decision-making on alternatives, post-project monitoring to test the predictive hypotheses, and a feed-back loop to add mitigation and refine predictions.*

- *Components neglected in the past include consideration of all alternatives such as no action, and the followup assessments and comparison with predictions.*

- *The predictive aspects of toxicity tests make them very suitable for use in EA.*

- *A technical expert presenting environmental information at a public hearing must make it clear and convincing, if it is to be useful in the decision-making process.*

This section gives a general outline of the major components of an environmental assessment. The scope then narrows to consider how toxicity testing fits into the scheme.

Environmental Assessment (EA) or Environmental Impact Assessment (EIA) is a process used to predict the probable effects a proposed human project could have on the natural environment and on human interests. It includes weighing of alternatives, and actions to avoid or lessen adverse affects (mitigation) or to remedy them. The process should include a formal decision on whether the project goes ahead, and monitoring afterwards to check if protective goals were achieved. Ecological Risk Assessment is a somewhat parallel process, but is limited to scientific and technical issues, and in fact, a risk assessment might be the scientific component of an EA. Risk assessment is described in Section 5.5.1.

At the national level, EAs are directed by the Canadian Environmental Assessment Agency under the Canadian Environmental Assessment Act, proclaimed in 1994. The act established a uniform basis for EA, including a principle of only one assessment for a given project, and incorporation of provisions for followup, to check accuracy of predictions and efficacy of mitigation. Provincial agencies carry out EAs with various legal bases.

5.4.1 The General Process

The initial stages of a complete EA are data-gathering and predictive processes. That is followed by decision-making, then a monitoring program with possible feedback to earlier stages (Figure 11).

A good EA parallels the scientific method in some of its features. It starts by defining the enterprise and the alternative approaches for accomplishing it (including one alternative of not doing it). Initial working hypotheses are developed for the likely effects of the alternative approaches (Figure 11; items 1, 2, 3). Programs are designed for gathering additional background information required (if any). There is a stage for formulation of modified or secondary hypotheses (item 5). Conclusions on the likely effects are part of the presentation to a reviewing body (item 8). After the project has proceeded, monitoring is followed by conclusions which will check the validity of hypotheses and predictions (items 14 and 15). In case of unsatisfactory findings, there is a recycling process for mitigation of the project, reformulation of hypotheses, and further monitoring (item 7).

An important but often neglected feature, is consideration of a wide range of alternatives. A project sponsor might have a single desired option, often the result of economic decisions. However, there are usually other ways of designing an industrial process, or laying out a housing development, and they should be considered. Cost-benefit analysis is implied at the initial stages of EA and at other stages. No-action alternatives should *always* be considered. What would be the economic and other effects of *not* discharging an effluent, or of *not* spraying with pesticide? What would be the effects of adopting *no* mitigating procedures? Such no-action alternatives serve as baselines for comparing alternatives and the benefits of remedial actions.

Failure to consider all alternatives is a common error in EA, although the principle has been stressed for decades (Leopold *et al.*, 1971; SCOPE, 1975). An early Environmental Assessment Act (Ontario, 1975) emphasized alternatives in six of nine basic steps of the process. Consideration of alternatives is particularly important when something is to be *done* (building something), or a chemical is to be used or replaced[51].

---

[51] The topic of alternatives is well described by SCOPE (1980).

"The weighting of alternatives is at the heart of the risk assessment process. In the control of chemical hazards, the evaluation of alternatives is particularly difficult because often the risks and benefits of the alternatives are as uncertain as the suspect chemical. In more recent decisions, the *risks of substitutes* which will inevitably fill the gap of the prohibited product, are taken more explicitly into account. Thus the proponents and opponents of the asbestos industry argue alike that the risks of substitutes such as glass fibre, which can have similar effects on human health to asbestos fibres, must be part of the overall assessment equation. At the same time, the additional deaths from fibres, etc., if asbestos is *not* used, needs to be set against the toll of ill health and premature death attributed to the use of asbestos. This view is put forward as a general principle for radiation protection:

The public must be protected from radiation but not to the extent that the degree of protection provided results in the substitution of a worse hazard for the radiation avoided. (BEIR, 1973)."

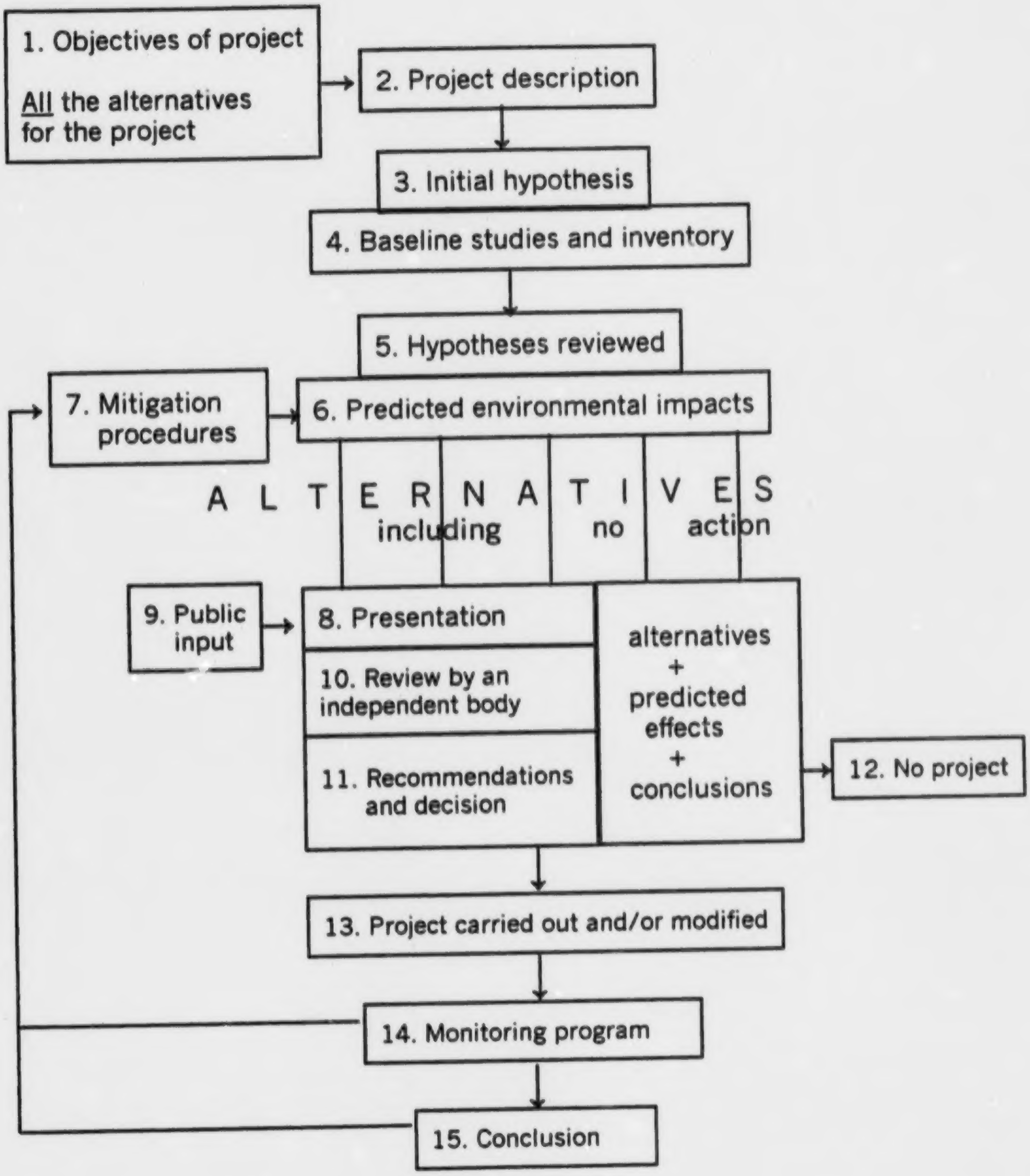
1. Objectives of project
All the alternatives for the project
2. Project description
3. Initial hypothesis
4. Baseline studies and inventory
5. Hypotheses reviewed
7. Mitigation procedures
6. Predicted environmental impacts
ALTERNATIVES
including no action
9. Public input
8. Presentation
10. Review by an independent body
11. Recommendations and decision
alternatives + predicted effects + conclusions
12. No project
13. Project carried out and/or modified
14. Monitoring program
15. Conclusion

Figure 11 **General Model of an Environmental Assessment.** The chart incorporates procedures and outlines from various sources.

Baseline surveys and inventories should be carefully designed to provide information that is directly required to make predictions and confirm or disprove hypotheses. Programs that set out to collect "all information possible" can lose sight of objectives and fail to answer the major questions about a development project.

Mitigation procedures (Figure 11; item 7) could range from minor items such as barriers against siltation of a stream, to major items such as movement of the entire project to another location.

EA is a public process (item 9), and it is important that the environmental case is clear and understandable, obviously credible, and convincing. Some suggestions on translating technical expertise into an effective presentation at a public hearing are given in Section 5.4.3.

Socio-economic benefits and effects are always major considerations in a project, but generally, environmental workers should not attempt to include those factors in their predictions or program. It is appropriate that environmental workers focus their efforts on ecological effects, and do not stray beyond their expertise. Attempts to place economic values on ecosystems or ecosystem components are unlikely to be successful. The major importance of a tree in an ecosystem might be reduced to the dollar value of the lumber which could be cut from it, if an economist's cost-benefit analysis were applied to the environment. It is more accurate, and more efficacious, to regard ecosystem protection as an absolute requirement.

Followup monitoring is vital in environmental assessments for aquatic systems (Larkin, 1984). Predictions might be incorrect, and there must be an effort to detect unexpected environmental damage. If detected, there must be a reactivation of the review process in EA (Figure 11). Criteria for environmental effects must have definite quantitative limits for what constitutes agreement with predictions (e.g., 95% probability that growth of vegetation will not be affected outside an effluent mixing zone). Three dimensions of effect should be measured and compared: (1) spatial (e.g., how far are deleterious effects detected?); (2) temporal (when do toxic episodes occur, and for how long?); and (3) severity (how strong are the effects?). Criteria for judging effect should be established before field work begins, and objective statistical methods should be used to accept or reject the hypotheses of environmental effects.

Some publications on EA are available in a Canadian context. An excellent critique by Beanlands and Duinker (1983) emphasizes good scientific practice and incorporating ecological principles. A book by Ramamoorthy and Baddaloo (1990) emphasizes environmental toxicology.

5.4.2 Toxicity Tests

Toxicity tests fit very well into EAs for projects that produce wastes which could leak, spill, or be discharged into soil or water surrounding the project. Toxicity tests are predictive and allow scientifically valid forecasts of ecological damage. After completion of the project, they combine with data from ecological surveys, to test the predictive hypotheses. *In-situ* tests might be useful (Section 4.14.1).

If the kinds and amounts of chemicals likely to reach the environment were known, predictions might use toxicity values obtained from the literature, or the substances could be tested. An effluent might even be simulated form its components. If the project were similar to existing industry, its toxicity could provide a first approximation. Bench-scale trials of treatment or mitigation processes could be assessed. The general principles in preceding parts of this document apply, in particular Section 4.2 is relevant, for laboratory-to-field predictions.

If one season or set of circumstances (e.g., low flow) could be identified as a worst-case situation, prediction of effects could adopt that situation. The frequency of worst-case conditions should be considered. Toxicity testing for field verification should normally use the same species and methods as in the predictive phase of the EA.

5.4.3 Public Presentation of Toxicity Information

Since public perception and public opinion has a role in EA, environmental and toxicological information should be presented in a fashion that makes it clear and useful to the public. Thorough treatments of communicating risks are available in the literature (USNRC, 1989; Leiss, 1989; and Kasperson and Stallen, 1991), but the following suggestions can provide some orientation.

- Present the information clearly and simply. Provide the key information in easily grasped charts or graphs, with only one kind of information per chart. Toxicity data superimposed on maps can be very effective. Summary text and comprehensible tabulations of data should also be available for those who want them.

- Avoid technical jargon and acronyms. Many of them can be expressed in words that will be understood! For example "96-h LC50 for rainbow trout" could be replaced with "lethal to an average trout".

- Certain words have an emotional loading in everyday parlance, and *toxic* is one; it is often taken as more or less synonymous with *poisonous* and *dangerous*. Some such terms can have substitutes which are less emotionally loaded, such as *biological effects* or *reduction in growth rate*. On the other hand, avoid euphemisms, which can create the impression of evading an unpalatable fact. A definition of toxicity and how it is measured might be in order. Cover in everyday terms, Paracelsus' explanation that everything is potentially toxic and that "the dose makes the poison". Household examples are useful, for example, orange juice can kill trout.

- Public perceptions are usually different from those of a specialist, and can be more concerned with broad implications than technical details. First, present the general meaning of the information for health of the ecosystem, then clarify details as required.

- A weight-of-evidence presentation can be convincing. Marshall information from surveys of indigenous species, environmental chemistry, toxicology, and parallel examples elsewhere.

- Put forward your conclusions concisely and dispassionately, and describe how the data support each conclusion. Keep the factually based conclusions separate from predictions and hypotheses. Explain why contrary evidence was considered less weighty.

- Stick with technical and scientific conclusions and opinions on environmental cause-and-effect. If you must voice social opinions on the project, try to save them all for after your technical items, and make it very clear that you are now speaking as a private citizen with personal preferences.

- Avoid patronizing your audience. Most people are experts in some subject in which you are not qualified, and most are capable of making rational decisions about matters that affect them. If people feel that a government or industry speaker is "talking down" to them, it will likely generate antagonism towards the speaker and to the information being presented.

5.5 Assessment and Remediation of Contaminated Sites

Key Guidance

- *Canada has a national program for remediation of contaminated sites (NCSRP), developed through the Canadian Council of Ministers of the Environment. Detailed guidance has been published by the Council.*

- ***Ecological Risk Assessment*** *(ERA) is a widely used predictive tool that is particularly relevant to the NCSRP. Several Canadian documents give detailed reviews of this topic. ERA has four steps. (1)* ***Receptor characterization*** *identifies the components of the ecosystem that might be affected by the toxic material(s). (2)* ***Hazard assessment*** *rates the toxic potency of the material(s). (3)* ***Exposure assessment*** *estimates the likely dose received by organisms. (4)* ***Risk characterization*** *combines the first three estimates, and estimates the probability and magnitude of real effects.*

- *During the initial assessment of the site, standard tests can be used to assess the toxicity of site components. A tiered approach is recommended, in conjunction with chemical measurements and biological surveys.*

- *In remediation, the biggest role of toxicity tests is for judging successful completion of the task, and monitoring its permanence.*

- *Biological guidelines of quality are being developed for sediments of the Great Lakes and other parts of Canada, in a procedure called BEAST. Physicochemical characteristics of the habitat are used to predict the macroinvertebrate community that would be likely in a clean situation, and also any potential toxicity of the sediment. Those predictions are compared with observed values to see if any added effects of pollution are evident. The British National Rivers Authority plans a further step, by setting biological objectives for water quality in rivers.*

- *Uncertainty is inherent in all steps of remediation. Conservative choices in each step can allow for that uncertainty, but might lead to "overly safe" conclusions. One way to avoid that is by probabilistic modelling which estimates the most likely scenario and also the likely range of outcomes with their associated probabilities.*

The Canadian Council of Ministers of Environment (CCME) established the National Contaminated Sites Remediation Program (NCSRP) in 1989 to promote a coordinated, nationally consistent approach to the identification, assessment and remediation of high-priority contaminated sites in Canada, especially "orphan" sites. More specific goals were to develop and demonstrate remediation technologies, and scientific tools for site assessment and remediation.

The goal of site remediation is to reduce the concentrations or bioavailability of toxicants to levels that are considered acceptable for human health and natural systems. A major element in this task is the establishment of objectives for quality. In the NCSRP, the objectives are chemical concentrations that are considered to produce no significant effects at the particular site in question. A set of interim quality guidelines for contaminated sites was adopted to address the urgent need (CCME, 1991a). Further information on guidelines for soil is given in Section 5.6.4.

Toxicity information is used in three general ways in the NCSRP. (1) Toxicity data are the base for developing environmental quality guidelines (see Section 5.6). The guidelines, such as the interim ones developed for the NCSRP (CCME, 1991a), are adapted to a specific site by considering the local setting, then used directly in assessing the site, and in monitoring. (2) Toxicity information is used in ecological risk assessment (see Sections 5.5.1 and 5.5.2). (3) Toxicity testing can be used directly for on-site investigations and monitoring, before, during, and after remediation (EC, 1996a).

Authoritative guidance documents are available under the NCSRP: quality guidelines (CCME, 1991a); a system for classifying sites (CCME, 1992); sampling and analysis (CCME, 1993b,c); the protocol for deriving soil quality guidelines (CCME, 1996a), and recommended soil quality guidelines (CCME, 1997a). Environment Canada reviews methods for toxicity testing of fresh water, sediments, and soil, with particular reference to contaminated sites (EC, 1994e).

5.5.1 Ecological Risk Assessment

Ecological Risk Assessment (ERA) is one of the tools that can be used in the NCSRP [52]. This section focuses on how environmental toxicology fits into risk assessment for the NCSRP. ERA has received much attention, and the approach is being described in major Canadian documents (CCME, 1996a,d,e; 1997a; EC, 1994e; 1996a; Gaudet *et al.*, 1995). The purpose of ERA, in the present context, is to assess the potential adverse effects of hazardous materials on ecosystems and their components, not excluding the interrelationships with human health. The ultimate goals of ERA in the context of site remediation, are first, to estimate the extent and nature of risk to the environment from the materials; and second, to determine the most effective strategy and goals for remediation (CCME, 1996d).

The approach for risk assessment at contaminated sites has four parts. **Problem formulation** is often included as an initial step, and might be added here, although it would seem to be a normal and required first step in any technical investigation, rather than a step that is distinctive for risk assessment.

1. **Receptor characterization** is aimed at identifying the components of the ecosystem (*receptors*, e.g., predatory birds) that could be affected by the toxic substances at a given site.

2. **Hazard assessment** is essentially a rating of the degree of danger associated with the material(s) in question. This stage evaluates the type and magnitude of adverse effects which can be caused by the material(s). This stage generally consists of examining information available for the material(s), identifying the nature of the hazards, formulating concentration-response relationships, and might include toxicity tests of contaminated media.

---

[52] The other main tool for assessing a site and developing a remediation program is to use regulations and quality guidelines that have been stated as chemical concentrations. An initial decision is made for a site, whether to use an approach that is criteria-based, or risk-based (explained in EC, 1994e). In a relatively straightforward situation, the method using criteria would suffice and a risk assessment need not be done. The procedures for the criteria-based method are not as difficult to comprehend and are not specifically outlined here; the general approach with quality guidelines is given in Section 5.6. The risk assessment could be either for human health or ecological aspects. The ecological approach is focussed upon herein.

3. **Exposure assessment** estimates the dose likely to be received by organisms. The process can be predictive, based on concentrations of toxicants at the site, and contact with organisms. This involves considering the potential routes of exposure for each toxicant and species, life cycles of organisms, and their seasonal movements. The estimate could also be retrospective, by examining body burdens or biochemical/physiological markers in exposed organisms.

4. **Risk characterization** is the process of estimating the probability and magnitude of real effects on organisms, by combining results of the exposure assessment and hazard assessment (EC, 1994e). The end result of the characterization, and hence of the ERA, is a statement about the risk of effects in various parts of the site, or estimates of the probability and magnitude of effects.

Ecological risk assessment is a complex and important topic for environmental toxicologists. It cannot be adequately covered here and the reader is advised to turn to other documents for specific guidance. A review by Environment Canada provides an astute and comprehensive introduction (EC, 1994e). The CCME and Environment Canada have been active in developing methods and frameworks for assessments, particularly for contaminated sites (CCME, 1996a,d,e; 1997a; EC, 1994e; 1996a; Gaudet *et al.*, 1995). The USEPA (1998) has formalized guidelines for conducting an assessment, in a document that will have strong influence within and outside the USA. Useful methods and reviews have also been published elsewhere (USEPA, 1989c; Suter, 1993; Calabrese and Baldwin, 1993). A succinct chapter by Suter (1997) assists in a clear understanding of the process. Approaches are described in two series of papers in issues of *Environmental Toxicology and Chemistry*, starting with editorial introductions by Bretthauer (1992), and Burns *et al.* (1994). Policy issues are considered in the proceedings of a workshop edited by de Peyster and Day (1998). Chapman *et al.* (1998) review ERA with emphasis on the safety factors involved. A book specifically on risk assessment for wetlands covers several case studies for these important habitats (Lewis *et al.*, 1999). The report from a workshop of OECD (1995b) is not a manual of procedures, but outlines areas of agreement of experts, mostly European, and areas where more research is needed. A brief outline of the topic is intended to explain the process to technical people who are not familiar with ERA, and could be useful for distributing to such people or groups (SETAC, 1997).

Investigators should remind themselves that there are certain assumptions inherent in ERA, assumptions that cannot be considered proven as fundamental truths. Indeed, some have been termed "fallacies" and "myths" by Power and McCarty (1997) who mention in particular: (1) the principle of basing assessments on data for a supposedly most sensitive species, with the assumption that the result will protect the community; (2) the variability of data for chronic and sublethal test results; and (3) the question of valid extrapolation from laboratory tests to a community in nature (see Section 4.1).

ERA is a formal procedure with distinct steps as outlined and described in reviews cited in the previous paragraph. Readers should be alert for very loose usage of the term in environmental literature of the 1990s. Unfortunately, there appears to be some degree of "band-wagon" use of the words *risk assessment* in titles of reports, published papers, and conferences. Sometimes the reports represent a complete ERA but often they describe instead, a toxicity test, a discrete development in mathematical modelling, or even an old-fashioned biological survey. (All these things can be part of ERA, but by themselves they do not constitute one.) Sometimes "risk assessment" is used to describe a process that is closer to environmental (impact) assessment, a term that seems to have lost favour. Environmental assessment is a more extensive set of procedures (Section 5.4), although it often includes ERA within its steps. The words "risk assessment" designate a specific set of procedures, and usage should be restricted to applications of the full procedure.

Some related topics might be mentioned. *Risk management* is a decision-making process that normally follows ERA. In considering what action should take place, particularly for remediation, risk management must include other factors such as technical feasibility, economics, and social, political, and legal constraints (Suter, 1995). This is part of environmental assessment (Section 5.4), and can become a *risk/benefit comparison* when alternatives are involved. The question of alternatives or options is mentioned for remediation in Section 5.5.3, and its importance is emphasized in Section 5.4.1.

5.5.2 Toxicity Tests for Assessing Contaminated Sites

Initial testing to investigate the extent of contamination at a site often uses an hierarchical approach. The first tier makes use of existing information (literature, surveys, etc.), but screening tests might be used to determine whether or not toxicity is present, and to determine the boundaries of contaminated areas. Environment Canada recommends simple, acute, lethal or sublethal tests for this phase (EC, 1994e).

In the second tier, more detailed testing might be used to delineate areas of high toxicity ("hot spots"), or elicit concentration-response relationships. This tier includes definitive testing with sensitive acute tests and sublethal tests. Because concentration-response information is important for the hazard assessment stage of an ERA, definitive second-tier tests could be multi-concentration ones, to determine LC50, EC50, IC25, etc.

Toxicity tests might be done on soil, sediment, porewater, surface water, or groundwater, depending on the site and historical knowledge of contamination. A testing program should suit the route of exposure. Soil toxicity is sometimes evaluated with elutriates or extracts and surface-water organisms (e.g., *Daphnia*), but it is better to use soil organisms (e.g., earthworms or terrestrial plants) unless there is a question of contaminated surface water or groundwater.

Batteries of several diverse toxicity tests are recommended for contaminated sites. The rationale was given in Section 2.3, but the major point is that different toxicants can affect different species. At least three species should be used, and Environment Canada recommends tests which are generally well known (Table 9).

Environment Canada (EC, 1996a) does not recommend departing from standard species and methods because the NCSRP is a national program, by definition, and benefits from comparable data. Additional tests on a local species such as in the seed germination test, however, might provide site-specific information. Similarly, it might sometimes be revealing to add tests at modified temperatures, pH values, or water content of soil. That could avoid stress on the organisms that was unrelated to the chemical contamination, and could reveal interactions between toxicity and site variables.

Toxicity testing on groundwater requires special consideration if the groundwater is remote from the surface biota. Nearby or distant ecosystems could be affected if the groundwater moved into the surface layers or emerged as a spring or seepage. Groundwater can undergo pronounced changes in pH, alkalinity, and ion balance as it migrates through soil or bedrock. Substances can be differentially adsorbed or desorbed according to their relative affinities for solid or aqueous phases. Consequently, the chemical profile of surfacing groundwater can be very different from that in samples taken at depth.

As always, information on toxicity is most effectively used when it is integrated with chemical investigation and biological surveys (Sections 1.2, 4.4). A combination of chemical profiles and toxicity tests can reveal the operative toxicants, an important step because bioavailability can vary widely, especially from partitioning to solids in soil and sediments (see Section 4.12.2 and 4.11.3). If there are a number of contaminants, toxicity tests with spiked clean soil can help identify the contribution of each. The procedures for toxicity identification can also be effective (TIE, Section 4.7).

Comparison of on-site monitoring of communities indicates whether toxic effects measured in the laboratory extend to the environment as expected, and demonstrate any local modifying factors. This comparison of field and laboratory data is, in fact, a part of the exposure assessment in ERA, and Environment Canada outlines this use of toxicity testing in more detail (EC, 1996a) [53].

If ERA is being used, a key step in hazard assessment is to determine the concentration-response relationships of toxic substances, to estimate threshold-effect concentrations (TOEC, sometimes termed "benchmarks" in this context). Alternatively, the threshold might be extrapolated using acute-to-chronic ratios (Section 4.2.2). Comparison of the estimated thresholds with established environmental quality guidelines could indicate whether the primary toxicants have been identified, since the guidelines are themselves estimates of no-effect concentrations.

---

[53] Although the role of field surveys of biota is not discussed in detail here, a strong case can be made for their use in ERA. Cairns and McCormick (1992) show the merits of a "top-down" retrospective approach, as a complement to predictive approaches. They cite examples worked out for past problems of contamination, e.g., predictive ERA could not have foretold the disastrous thinning of bird eggshells from accumulating pesticides.

Table 9 Toxicity Tests Recommended for the National Contaminated Sites Remediation Program (slightly modified from EC, 1996a)

Medium	Organisms	Endpoints	Type	Source
Soil	Lettuce and radish seeds *(Lactuca sativa, Raphanus sativa)*	seed germination and seedling emergence	sublethal, acute	Greene *et al.*, 1989
	Earthworm *(Eisenia andrei)*	mortality	lethal, acute	Greene *et al.*, 1989
	Alga *(Selenastrum capricornutum)*	growth inhibition	sublethal, chronic	EC, 1992d
Freshwater sediment	Amphipod *(Hyalella azteca)*	mortality sexual maturation	lethal & sublethal, chronic	ASTM, 1997a
	Midge *(Chironomus tentans)*	mortality	lethal, chronic	Bedard *et al.*, 1992
	Mayfly (*Hexagenia* spp.)	mortality	lethal, acute	Bedard *et al.*, 1992
	Alga *(Selenastrum capricornutum)*	growth inhibition	chronic, sublethal	EC, 1992d
Fresh water	Bacterium *(Vibrio fischeri*, formerly *Photobacterium phosphoreum*)	luminescence inhibition	acute, sublethal	EC, 1992c
	Alga *(Selenastrum capricornutum)*	growth inhibition	chronic, sublethal	EC, 1992d
	Cladoceran *(Daphnia magna, D. pulex)* *(Ceriodaphnia dubia)*	mortality or mortality and reproduction	acute or chronic, lethal & sublethal	EC, 1990c; 1992a
	Rainbow trout *(Onchorynchus mykiss)*	mortality of fry	acute, lethal	EC, 1990a
	Fathead minnow *(Pimephales promelas)*	mortality and growth of larvae	acute, sublethal	EC, 1992b

The final step of an ERA combines the hazard assessment (derived from toxicity information) with the exposure assessment, to produce a risk characterization for the site. The risk characterization should identify the ecosystem components most at risk, and the primary toxicants and pathways. A successful characterization provides an objective means to rank the site for urgency of remediation, and for priority compared to other sites.

5.5.3 Toxicity Testing in Remediation of Sites

The alternatives available for remediation *must* always be given due consideration, and risk assessment can be extended to help decide among the options (SCOPE, 1975; 1980). Remediation involves some degree of disturbance such as vehicular traffic, excavation of soil, drilling wells, erecting barriers, or adding nutrients, chelators, etc. The disturbances create some ecological hazard, and might change pathways of the toxicants. At some sites, the risks of remediation might be greater than the risk of leaving the site alone (e.g., a harbour with contaminated sediments). Other aspects of considering alternatives are mentioned under Environmental Assessment in Section 5.4.1.

Once a decision has been made to proceed with remediation, toxicity testing could be used to monitor the progress. The thresholds of toxic response defined during assessment provide relevant targets. Toxicity tests could confirm that attaining the chemical targets had actually removed the toxicity of the site materials, and this combination of approaches would be especially rewarding in cases with a mixture of chemicals, for which prediction is difficult. With parallel data, the decline in toxicity should be correlated with the declining concentrations of the substance(s) thought to be responsible for the toxic effects. If that did not occur, remediation might be focusing on the wrong contaminants. Targets should be re-evaluated before incurring further cleanup costs.

At the conclusion of remediation, the site should be thoroughly surveyed once more for chemical levels, toxicity, and communities of organisms, to confirm that risk has declined to a tolerable level. Confirmatory testing should ideally employ a full battery of toxicity tests, although in straightforward situations it could be satisfactory to use the test that was shown to be most sensitive in the preceding stages. Meier *et al.* (1997) provide an example of using soil toxicity tests to evaluate the success of remediation at a site which was treated to remove PCBs. They made use of two tests with earthworms, one a 14-d test measuring mortality, and the other a 21-d test measuring reproduction. Seed germination and root elongation were assessed in lettuce and oats, and onions were used for genotoxicity (anaphase aberrations).

Environment Canada recommends two rounds of confirmatory testing (EC, 1996a). The early round would be immediately after completion of remediation, with the purposes indicated previously. A second round over a longer term would indicate whether satisfactory levels of contaminants were continuing. Objectives of long-term monitoring *must* be clearly established before it begins, and must include definite decision criteria that would trigger management decisions; without those criteria, monitoring could be endless and pointless (EC, 1996a). A guidance manual is available for using toxicity tests for site-specific remediation criteria for Canadian oil and gas facilities (Golder, 1996).

5.5.4 Biological Guidelines for Remediation of Contaminated Sediment

Parallel chemical and biological approaches are evident in Ontario for evaluating contaminated sediments. The province has a set of chemical guidelines for quality of aquatic sediment (Persaud *et al.*, 1992). In a biological approach, researchers from Environment Canada (Ontario Region) have developed a system combining benthic community analyses and toxicity tests on sediments (Reynoldson *et al.*, 1995). The *Benthic Assessment of Sediment* method *(BEAST)* is intended for setting quality guidelines for sediments of the nearshore areas of the Laurentian Great Lakes. Reynoldson *et al.* (1995) expect to establish remediation targets for sediments in each "Area of Concern" on the Canadian side of the Great Lakes. The system has "demonstrated the ability to predict the community structure of benthic invertebrates at reference sites using simple habitat and water quality descriptors and multivariate analyses" (Day *et al.*, 1996).

The BEAST method uses multivariate statistical methods to relate the structure of benthic invertebrate communities at uncontaminated sites, to simple variables describing quality of sediment and water. The relationships are then used to predict the potential invertebrate communities at contaminated sites, *if the sites were clean.* The predicted communities are compared with those actually found.

The same methods relate toxicity findings for each uncontaminated sediment with physicochemical variables. If a sediment deviates from the predictions for the invertebrate community or toxicity, contamination is suspected. Hence, site-specific biological objectives can be set on the basis of measured physicochemical characteristics of the habitat, and conclusions can also be drawn on the apparent anthropogenic degradation.

A data base was assembled from 346 reference sites along Lake Michigan and the Canadian side of the other Great Lakes and included three major kinds of information.

1. The structure of the benthic invertebrate communities.

2. Key variables of the habitat.

3. The responses to sediment from the same sites, shown by four species of benthic invertebrates *(Hyalella azteca, Chironomus riparius, Hexagenia* spp. and *Tubifex)*. The toxicity tests were included to confirm that the sediment was the cause of observed effects in the field, rather than the water column or other physical disturbances.

Results for 96 reference sites are available, and the sampling program continues (Reynoldson *et al.*, 1995). Twenty-seven, non-pollutional environmental variables have been included in analyses. They are broad descriptors of sediment (e.g., major elements, particle size, and organic material), and general attributes of the site (e.g., depth and water chemistry).

Models were developed from results at 96 sites to predict benthic community structure in clean water, and toxicity-test responses. The best predictors of community structure were water depth, sediment type, alkalinity, and trophic status. The model successfully classified 87% of the sites into one of five classes. Toxicity-test responses were less easily predicted and the best model contained some 19 variables, embracing alkalinity, major ions, nutrients, organic matter content, particle size, and depth. The toxicity model correctly classified 70% of the sites.

The utility of the BEAST approach was demonstrated by an assessment of remediation needs in Collingwood Harbour in Georgian Bay, Lake Huron, identified as an area of concern. The harbour was heavily contaminated with metals (copper, zinc, lead, arsenic, and iron) according to Ontario guidelines (Persaud *et al.*, 1992), and sediment removal was being considered. The community composition predicted by BEAST at most of 25 points in the harbour was comparable to that actually observed, indicating no significant impairment. At most sites, observed toxicity was within the predicted range for clean sites, except for a few places near shore with very high concentrations of metals. In short, the BEAST method indicated that costly dredging was unnecessary, except for a few key areas of the inner harbour (Reynoldson *et al.*, 1995).

The BEAST technique awaits completion of the sampling program. The approach has obvious merits, but also some limitations. It is labour intensive, and requires an extensive data base to satisfy nonparametric statistical methods. The resulting model cannot be applied outside the range of variation in reference sites, and the effects of seasonal and yearly variation are still uncertain.

A British system of biotic guidelines. Techniques developed in the United Kingdom were forerunners of parts of the Canadian BEAST method. The British system does not use toxicity data to develop quality guidelines, but provides an instructive example of biological guidelines. UK researchers developed methods to predict the macroinvertebrate fauna of clean streams, based on physicochemical measurements of the habitats (gradient, velocity, substrate, mineral content of water, etc.; Wright *et al.*, 1984; Moss *et al.*, 1987; Armitage *et al.*, 1987). The physicochemical measurements from a given riverine location can now be entered into a computer model (RIVPACS) to predict the expected clean-water community with good dependability.

Britain has also developed numerical biotic indices of water pollution based on the results of biological surveys of benthic invertebrates. Over the decades, the biotic indices have become streamlined, simplified, and easy to use. The predictive computer program was put together with the biotic index by the National Rivers Authority (NRA) which is in charge of river water quality for England and Wales (NRA, 1991).

The NRA appears to be proceeding towards use of the biological system in parallel with chemical water quality objectives, for three aspects of water quality management. First, a river is classified according to water quality, partly on the basis of the biotic index and partly on chemical and bacterial guidelines. Second, a selected value of the biotic index is used as a water quality *objective* for a section of river. Finally, the indices obtained in later monitoring, are used to check compliance with the objectives.

The UK agency has gone further by installing a "biological override" in the procedure, in case the biotic and chemical approaches disagree in classifying a section of river. The biotic information would take precedence. When tested on real information, the biological override had a strong effect of upgrading the classification of sections of rivers, compared to classification based on chemical water quality objectives (NRA, 1991).

5.5.5 Allowing for Uncertainty

Every stage of the site assessment and remediation process must consider the uncertainty associated with estimates. Uncertainty is present in the concept of risk itself, by definition (Suter, 1990). Contradictory results from different assessment approaches are inevitable (Burmaster *et al.*, 1991).

The first source of uncertainty which can be distinguished in use of toxicity tests arises from inaccuracy and imprecision of instruments, measurements, test procedures, and analysis of results. That is similar to uncertainty with other techniques such as chemical analyses, represents controllable error, and should be reduced as far as possible. The uncertainty about endpoints such as the LC50 is expressed as 95% confidence limits, which should be included with every definitive toxicity test. The uncertainty from variability of testing methods is retained when the results are extrapolated to effects in the environment.

The second source of uncertainty is random variation in natural processes, which cannot be eliminated and must be treated as a variable in the information base.

Uncertainty is an integral component of ERA, so incorrect estimates of hazard or exposure could lead to erroneously high or low estimates of risk. Underestimating risk could have serious consequences for the site and its ecosystem. The usual approach is to adopt a conservative stance, for example, by choosing a low estimate of toxic threshold, if several are available.

A better solution is to incorporate risk explicitly by adopting probabilistic modelling. Results at each stage of the ERA are expressed as a set of possibilities, each with an attached estimate of probability. The most sophisticated variants of this idea use stochastic simulation models, which make use of random choices. The input data are not fixed values, but distributions based on the range of field and laboratory data. Each run of the model randomly chooses one value from the range of possibilities for each input variable. The simulation is repeated many times. All the output estimates are combined into a probability distribution of possible outcomes, which indicates the most likely events and the range of events that might occur.

The advantages of a probabilistic approach are several. The uncertainty in each aspect of the data is considered and more realistic estimates of ecological risk are provided. Probabilistic simulations allow statements on the possible outcomes and the likelihood of each, a benefit for site management. The conservatism inherent in fixed-parameter approaches is avoided. The model can also guide analysis by indicating which input parameters cause the greatest change in the output (sensitivity analysis). Sampling and analysis can be more efficient by redirecting effort toward those more crucial items.

This discussion has centred on ERA, but the same principle applies to more qualitative assessments. The best approach to allow for uncertainty is to measure the same variables using diverse methods, and to assign the greatest credibility to observations that are confirmed by more than one method. As frequently mentioned in this document, one way of doing that is to compare toxicity results with biological surveys and environmental chemistry. For more information on the special problems of uncertainty in ecological risk assessment, see EC (1994e), Suter (1990; 1993) and the proceedings of a workshop devoted entirely to this topic (Warren-Hicks and Moore, 1998).

5.6 Development of Environmental Quality Guidelines

Key Guidance

- *The Canadian Council of Ministers of the Environment (CCME) is responsible for setting national guidelines for environmental quality. There are also provincial and regional guidelines.*

- *A Canadian guideline for environmental quality of a given substance is usually a single maximum concentration, intended to protect all kinds and life stages of organisms during indefinite exposure. Total concentration is normally used, although a chemical fraction might be specified if it could be done with certainty.*

- *To define a water quality guideline for aquatic life, information from the literature is first graded for acceptability. A full guideline requires at least three good estimates of toxicity for fishes, two for invertebrates, one for plants, and several other kinds of physicochemical and biotic information. The guideline is 1/10 of the lowest LOEC for a Canadian species.*

- *For irrigation water used in agriculture, guidelines are similarly derived from long-term studies of toxicity to crop plants, using concentration in the water for pesticides, or in the soil for other chemicals. Factors of 0.1 are applied to the lowest TOECs to derive guidelines for (a) cereals, hay, and pasture, and (b) other crops.*

- *An interim guideline for sediments requires that a similar compilation and screening of data must yield 20 tests that show toxicity, and 20 that do not. The distributions are used to calculate a guideline. A safety factor might be used for high uncertainty of data. A full guideline requires further information associating type of sediment with the toxicity results. The approach can be either "weight-of-evidence" or tests with spiked sediments, but routine use of the spiked tests awaits standard methods.*

- *Interim quality guidelines for soil were compiled from those existing elsewhere, to satisfy needs of the National Contaminated Sites Remediation Program. There is now a protocol for deriving soil guidelines, which fits the general CCME pattern. For each substance, separate guidelines can be derived for four categories of land use: agricultural, residential and parks, commercial, and industrial.*

- *Environmental quality objectives can be derived from guidelines, to account for local species, seasonal changes, long-term effects, and stress from other agents.*

There are distinct forms of quality limits set for environmental protection.

- A *guideline* is a scientifically based numerical concentration limit, or narrative statement, recommended to support and maintain a designated use of soil, sediment, or water. (In jurisdictions such as the USA, and some European agencies, *criterion* is used in the sense of *guideline*.)

- An *objective* can receive the same definition as guideline, except that it has been established to protect and maintain a specified use of soil, sediment, or water at a particular location by taking into account site-specific conditions. Some provinces have established lists of water quality objectives, and they reflect "officially desired conditions".

- A *standard* is similar to an objective except that it is recognized in enforceable environmental control laws of a level of government.

5.6.1 Background

Setting national environmental quality guidelines is the responsibility of the Canadian Council of Ministers of the Environment (CCME) and its predecessor, the Canadian Council of Resource and Environment Ministers (CCREM). National guidelines for protection of aquatic life, and use of water for agriculture (irrigation and livestock watering), drinking, industry, and recreation and aesthetics have been published (CCREM, 1987; CCME, 1998). There is a protocol for developing guidelines for residues of substances in tissues, with draft guidelines for dioxins and furans (CCME, 1996c; 1997b). Protocols for deriving guidelines for soils, contaminated sites, and sediments have been finalized (CCME, 1995; 1996a), while interim or definitive quality guidelines are available (CCME, 1991a; 1997a; EC, 1995c).

The provinces have produced environmental quality guidelines or objectives for water (e.g., AEP, 1978; OMOE, 1984; B.C. MELP, 1998) or sediment (Persaud *et al.*, 1992). Ontario has described its process for deriving objectives (OMOE, 1991). Other jurisdictions such as the International Joint Commission (IJC) and the Prairie Provinces Water Board (PPWB) have guidelines or objectives. The approaches of Canadian agencies are reviewed by Taylor (1989).

National guidelines are intended as conservative estimates of toxicant concentrations that will not cause adverse biological effects in any of the great variety of biomes and ecosystems across the country. They provide strong guidance for suitable physicochemical conditions in the environment, and can be used in several ways. In surveys, they can be used to judge whether environmental conditions appear to be satisfactory. For monitoring or regulatory purposes, they can serve as a basis for deriving objectives or standards. They are particularly useful for predictive calculations (Sections 1.2, 4.4). For example, the measured concentration of a toxicant in an effluent can be extrapolated to the receiving water, allowing for dilution, to predict whether a guideline would be exceeded. At the design stage, presumed characteristics of an effluent could be extrapolated to the receiving water, to see if guidelines would be met.

This section provides a brief overview of how toxicity data are used to develop environmental quality guidelines by the CCME. The procedures in place for deriving guidelines for water, sediment, and soil quality are described.

5.6.2 Water Quality Guidelines

Guidelines for protection of aquatic life and crops (irrigation water) are discussed here. The protocol for developing national water quality guidelines for protection of aquatic life is given in CCREM (1987), and in an update with improvements for selecting data and deriving guidelines (CCME, 1991b).

For **aquatic life**, the revised protocol is founded on four basic principles.

1. The main groups of larger aquatic organisms are considered if data are available (algae, macrophytes, invertebrates, fish, amphibians). For limited data, interim guidelines are preferred over no guidelines.

2. Guidelines are set at levels intended to protect all forms and stages of aquatic life during an indefinite exposure. The guidelines are derived on a scientific basis, without consideration of whether they can be achieved through management.

3. For most water quality variables, a single maximum value is recommended as the national guideline.

4. Guidelines normally refer to the total concentration of a chemical in an unfiltered sample. Guidelines can be based on some fraction of the total concentration, only if toxicity can be firmly assigned to that fraction, and analytical techniques can routinely and unequivocally identify the fraction.

The protocol for deriving Canadian water quality guidelines has four major steps (Figure 12). (The methods for developing guidelines in other media have similar steps.)

a. The literature is searched to obtain information on physical and chemical properties, environmental concentrations, fates and behaviours, bioaccumulation potential, acute and chronic toxicity, genotoxicity, and information from other jurisdictions concerning the chemical under review. Environmental data must be sufficient to identify major pathways and fates of the chemical in the aquatic environment.

b. Minimum requirements must be met for acceptance of data on toxicity (Table 10) and on environmental fate. Interim guidelines can be derived from sets of data that do not meet the requirements. Exceptions to the requirements in Table 10 can be allowed for justifiable scientific reasons, e.g., reliable acute-to-chronic ratios can substitute for sublethal tests. The major environmental pathways and fates of the chemical in the aquatic environment must be identified.

c. Each toxicity report is classified as *primary* (reliable flow-through tests using standard methods and measured toxicant concentrations), *secondary* (reliable static tests including those using modified methods), or *unacceptable*. The minimum data set must be complete and composed of primary tests, to proceed with derivation of a guideline. Primary or secondary data may be used for interim guidelines.

d. The single lowest LOEC from a sublethal and chronic study on a native Canadian species is multiplied by 0.1 to arrive at the final guideline. The multiplication by 0.1 represents the application of a safety factor, and is intended to allow for differences in sensitivity among untested species and endpoints, and for extrapolation to the environment. If applicable sublethal and chronic data are not available, the guideline is derived by dividing the lowest primary acute LC50 or EC50 by an acute-to-chronic ratio, or failing that, multiplying by a universal application factor. The application factor is 0.05 for nonpersistent chemicals (half-life in water <8 wk), and 0.01 for persistent chemicals. (See Section 4.2.1, 4.2.2 for acute-to-chronic ratios and application factors.)

For **irrigation water**, the Canadian protocol is intended to derive quality guidelines that protect crops from repeated

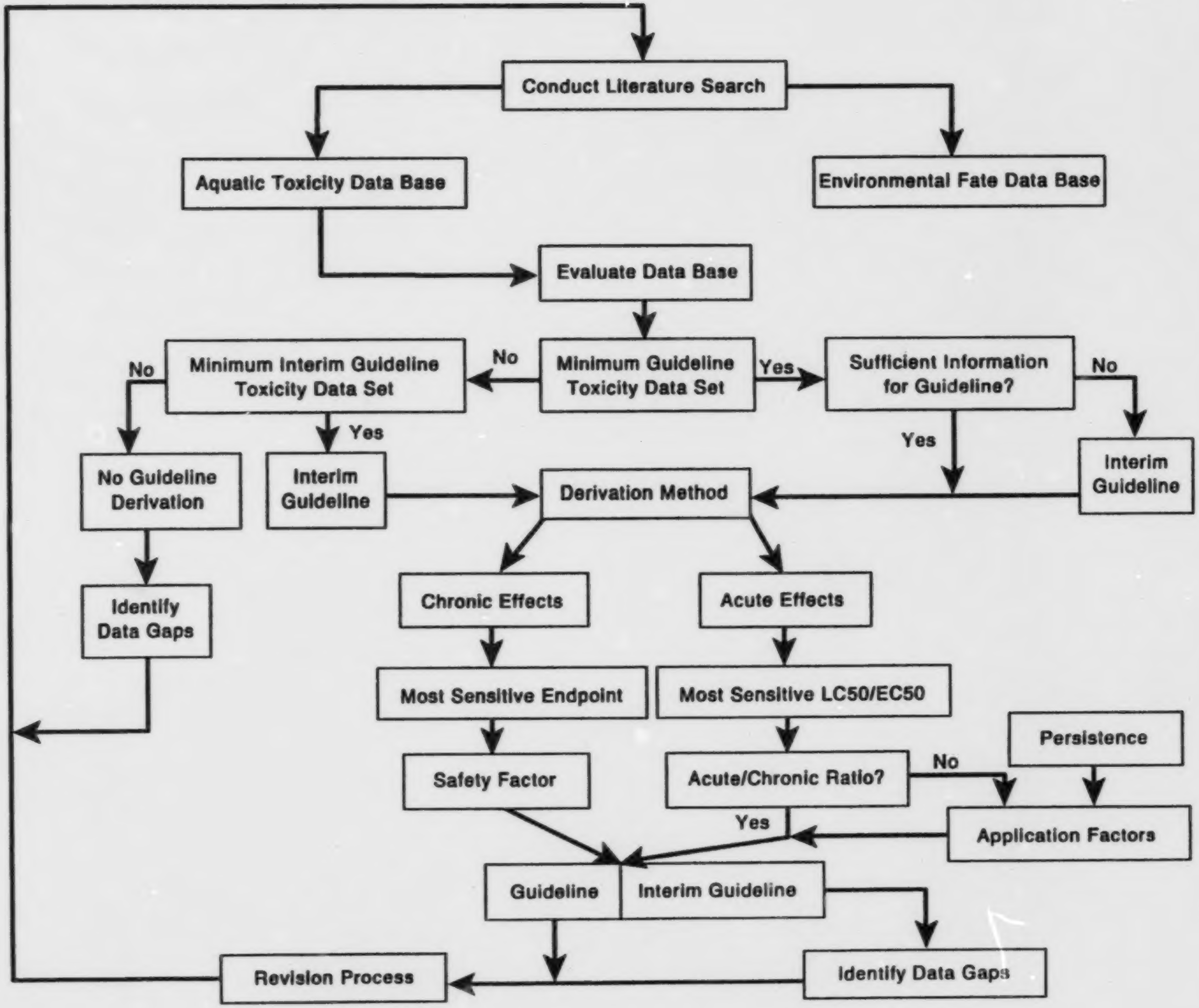
Conduct Literature Search
Aquatic Toxicity Data Base
Environmental Fate Data Base
Evaluate Data Base
Minimum Interim Guideline Toxicity Data Set
No
Minimum Guideline Toxicity Data Set
Yes
Sufficient Information for Guideline?
No
Yes
No Guideline Derivation
Interim Guideline
Derivation Method
Yes
Interim Guideline
Identify Data Gaps
Chronic Effects
Acute Effects
Most Sensitive Endpoint
Most Sensitive LC50/EC50
Persistence
Safety Factor
Acute/Chronic Ratio?
No
Yes
Application Factors
Guideline
Interim Guideline
Revision Process
Identify Data Gaps

Figure 12 The Protocol of the Canadian Council of Ministers of the Environment for Deriving Water Quality Guidelines for the Protection of Aquatic Organisms (from CCME, 1991b)

Table 10 **The Minimum Set of Toxicity Tests Required for Deriving a Canadian Water Quality Guideline for Aquatic Life (from CCME, 1991b)**

---

1. Freshwater guidelines

Fish	• at least three studies on three or more freshwater species resident in North America, including at least one cold-water species (e.g., trout) and one warm-water species (e.g., fathead minnow) • of these studies, ≥2 must be chronic* studies
Invertebrates	• at least two chronic* studies on ≥2 invertebrates from different classes, one of which is a planktonic species resident in North America (e.g., daphnid)
Plants	• at least one study on a freshwater vascular plant or alga resident in North America • for highly phytotoxic chemicals, four acute or chronic studies on non-target freshwater plants or algae

2. Interim freshwater guidelines

Fish	• at least two acute or chronic studies on ≥2 fishes, one of which is a cold-water species (e.g., trout) resident in North America
Invertebrates	• at least two acute or chronic studies on ≥2 invertebrates from different classes, one of which is a planktonic species resident in North America (e.g., daphnid)

3. Marine guidelines

Fish	• at least three studies on ≥3 temperate marine fishes, including ≥2 chronic* studies
Invertebrates	• at least two chronic* studies on ≥2 temperate marine invertebrates from different classes
Plants	• at least one study on a temperate marine vascular plant or marine alga

4. Interim marine guidelines

Fish	• at least two acute or chronic studies on ≥2 marine fishes, one of which is a temperate species
Invertebrates	• at least two acute or chronic studies on ≥2 marine species from different classes, one of which is a temperate species

---

* In these requirements of CCME, the word *chronic* is used to signify a full life-cycle or partial life-cycle study, with observations on sublethal effects.

applications of water over a full growing season (CCME, 1993a). The protocol follows the same principles outlined previously for aquatic life, modified to fit crops. Guidelines are based on two major pieces of information: (1) maximum irrigation rates in L/ha·y; and (2) sensitivities of crops, measured for pesticides as acceptable application rates in kg of active ingredient (a.i.) per ha [kg (a.i.)/ha], and for industrial or other chemicals by acceptable soil concentrations in mg/kg of soil.

Irrigation guidelines are derived from dose-response data for sensitive crops grown in Canada, preferably from long-term studies with irrigation water. The method calculates a maximum acceptable toxicant concentration for a species (SMATC) by multiplying the TOEC by 0.1. If there are too few studies using irrigation water, the TOEC from other chronic studies is multiplied by 0.1 to derive an *acceptable application rate* (kg/ha) for a pesticide or an *acceptable soil concentration* (mg/kg) for other chemicals. Those are estimates of levels that will not cause adverse effects over one growing season. The acceptable application rate for pesticides is divided by the maximum annual irrigation rate for the crop in Canada, to obtain the SMATC in µg/L. For other chemicals, the acceptable soil concentration is multiplied by an estimate of the mass of 1 ha of soil to rooting depth, and that is divided by the maximum annual irrigation rate to obtain the SMATC.

All the available SMATCs for a given chemical, however derived, are compiled in two groups: those for cereals, tame hay, and pasture; and those for other crops. An irrigation water quality guideline is derived for each group by selecting the lowest SMATC for the group.

5.6.3 Sediment Quality Guidelines

Each of several methods for deriving sediment quality guidelines has advantages and disadvantages, and these are appraised in comprehensive fashion by MacDonald *et al.* (1992), who also compiled the quality assessment values from international sources. Approaches are also reviewed by Chapman (1989b) and Persaud *et al.* (1992).

The procedure for deriving Canadian sediment guidelines (CCME, 1995; Smith *et al.*, 1996a) calls for complementary use of two approaches: a "weight-of-evidence" approach developed in the USA for the *National Status and Trends Program* (Long and Morgan, 1990), and toxicity tests with spiked sediments. Ideally, sediment guidelines would be derived from detailed concentration-response relationships, tests which can only be run by spiking sediments (see Section 4.11.1). That would also allow control of modifying factors and testing of sensitive life stages. A protocol for deriving guidelines from spiked-sediment tests has been developed (CCME, 1995), but will rarely be used in the near term because the methods for such tests have not been standardized, and because data are limited. Until those uncertainties have been resolved, use will be made of a modified version of the approach from the National Status and Trends Program. It involves evaluation of various chemical and biological data: models of equilibrium partitioning; assessment values from other jurisdictions; spiked-sediment toxicity tests; and field studies (CCME, 1995). Threshold-effect concentrations of chemicals in sediment are calculated, and recommended as a sediment quality guideline or interim guideline.

The four guiding principles for deriving sediment guidelines are similar to those for water, and the first two are essentially identical (see Section 5.6.2; items 1 and 2). The first principle is that all kinds of organisms are considered if data are available. The second is that guidelines are intended to protect all kinds and stages of organisms. Thirdly, guidelines normally refer to total concentration of a substance in surficial sediments on a dry weight basis, but guidelines will be suitably revised when the influence of a modifying factor is understood (e.g., total organic carbon modifying toxicity of nonpolar organics). The fourth principle reiterates that guidelines will be reviewed and amended periodically, in light of new scientific information.

The protocol for sediment guidelines also parallels that for water. Marine and freshwater guidelines are developed separately, according to the following sequence (CCME, 1995).

a. The literature is searched to obtain information on physicochemical properties, fate, and toxicity, as for water guidelines (Section 5.6.2, item a).

b. Toxicological data are evaluated for quality, as for water (Section 5.6.2, item b).

c. All applicable North American data for the substance are compiled, and sorted in ascending order of concentration to produce a table. Each row in the table consists of the measured chemical concentration, location, analysis type or approach, test duration, endpoint measured, species and life stage tested, whether associated biological effects were observed, and the study reference. Data on characteristics of the sediment and overlying water are also summarized, where available.

d. Minimum requirements for data on toxicity and environmental fata must be met or the process stops. To derive an interim sediment guideline, the data for toxicity must include at least 20 entries associated with biological effects and another 20 which are not. To derive a full guideline from an interim guideline, sufficient supporting information must be available to link the interim one with specific types of sediment and/or characteristics of the sediment or the overlying water. Derivation of interim guidelines can use certain other processes depending on available data, but preference is given to biological effects.

e. If the minimum requirements for data are met for an interim sediment quality guideline, the threshold effects level is calculated as the geometric mean of the lower 15th percentile concentration in the Effects data and the 50th percentile concentration of the No-Effects data. The threshold effects level is designed as the concentration below which adverse biological effects are almost never observed.

f. If a high degree of uncertainty is associated with the threshold effects level, a safety factor might be applied. The rationale for choosing a factor and the uncertainties which might be relevant are provided in CCME (1995). Otherwise, the threshold effects level is recommended as the interim sediment quality guideline.

Smith *et al.* (1996a) provide a background on how the technical data were collected, and the rationale for their use. Analysis of the data accumulated for each chemical provided the above-mentioned threshold effect level (TEL), and also the probable effect level (PEL), above which adverse effects were expected to be frequent. The data supported calculation of TELs and PELs for 31 substances in marine sediments, including nine metals, 13 PAHs, total PCBs, seven pesticides and one phthalate ester. For freshwater sediments, levels were calculated for eight metals, six PAHs, total PCBs, and eight pesticides. The TELs were used as draft interim sediment quality guidelines.

As with all chemical guidelines, these are only predictions of toxicity. Smith *et al.* (1996a) point out that "the predictive capability of the guidelines should be evaluated through field validations at a variety of sites across Canada", that they should be used with other complementary tools such as toxicity tests, and that the guidelines are not intended to serve directly as site-specific objectives without consideration of local management factors. Internal evaluation (within the same data sets) indicated that the TELs had relatively good reliability, with less than 10% incidence of adverse effects at concentrations below the TEL (Smith *et al.* 1996b). The PELs appeared less reliable, because higher concentrations were without , adverse effects in more than half of the cases, for metals PAHs, and two pesticides. Reliability of the PELs for three other pesticides was good, with adverse effects in about 70–85% of the cases with higher concentrations *(loc. cit.)*. O'Connor *et al.* (1998) compared toxicity and chemical characteristics for sediments from a number of locations around the USA, starting with the proposition: "Theoretical reasons exist for doubting that any of these chemical-based guidelines reflect sediment toxicity". Their comparison of data provided more encouraging conclusions than that, but they recommended that the guidelines "should not be used alone as indicators of toxicity". They considered "that guidelines based on bulk chemistry can provide useful triggers for further analysis ...", advice that corresponds to good modern practice of using multiple lines of evidence.

5.6.4 Soil Quality Guidelines

Guidelines for soils have developed in close association with activities for contaminated sites, and relevant information is given in Section 5.5. Interim quality guidelines were adopted to address an urgent need of the National Contaminated Sites Remediation Program (CCME, 1991a). These were compiled from existing guidelines in various Canadian jurisdictions. Interim guidelines for organic substances were largely based on analytical detection limits, while those for inorganic substances were mainly based on background concentrations. It was recognized that the interim guidelines lacked supporting scientific rationale and would be replaced by effects-based guidelines once a protocol was available, along with sufficient toxicological data for terrestrial organisms.

The protocol developed (CCME, 1996a) provides the rationale and procedures to derive guidelines for four categories of land use (agricultural, residential/parkland, commercial and industrial). Environmental and human health guidelines are developed in parallel, and the lower of the two is selected as the final soil quality guideline. The human health assessment does not rely on ecotoxicological data, so it is not further discussed here, but details are given in the protocol.

Effects-based soil quality guidelines indicate concentrations at which no adverse effects are anticipated on either human health or the natural community. Comparison with the guidelines for a specified land use can indicate the need for further investigation or remediation, and the scale and urgency of further action. Soil quality guidelines, like other guidelines, can provide a basis for objectives at a given locality or site. If remediation is not feasible for technological or other reasons, the guidelines can also provide guidance on the need for land-use restrictions or other management options to protect human health and the environment.

Soil is a complex and heterogeneous medium. A variety of organisms from microorganisms to vertebrates can be exposed to soil toxicants through many pathways, including direct contact, porewater, ingestion of soil or contaminated food, and inhalation of contaminated gas. The procedure for deriving a guideline for soil quality considers that complexity. It also follows standard practice developed by the CCME, and is therefore similar to that outlined for water (Section 5.6.2).

- The literature is reviewed to obtain information on physicochemical properties and toxicity, as listed for water, including information from other jurisdictions.
- All toxicological reports are evaluated for scientific quality before being used, and are considered unacceptable if they have inadequate test design, procedures, or analysis, or if insufficient information was reported to evaluate those factors.
- There are minimum data requirements for the three methods of deriving guidelines: the preferred one uses *weight of evidence*; the second is based on *LOECs*, and the third uses *median effects*. The choice of method depends on the type and amount of acceptable toxicological data. In general, at a minimum, data are required on toxicity to terrestrial invertebrates and terrestrial plants. To ensure the protection of microbial processes, the resulting soil contact number is evaluated for microbial cycling of nutrients and energy, and adjusted if required. For agricultural land use, an added procedure derives a soil quality guideline based on ingestion of food and soil by livestock and wildlife.

A groundwater protection check identifies the soil quality needed to maintain groundwater fit for human consumption. The concentration of organic substances in groundwater is generally governed by an equilibrium partitioning relationship which depends on the level of organic matter in the soil and the tendency of the compound to bind with that organic matter. This equilibrium is taken into account for guidelines for organic substances.

Guideline derivation considers key receptors and exposure pathways for four land uses (Table 11).

For ***agricultural land use***, the guidelines seek to protect the primary activities of growing crops and producing livestock. Protection must also be offered to resident and transitory wildlife and native flora because they may use agricultural land.

The ***residential/parkland*** category is primarily designated for human housing or recreation. It may include campgrounds, but does not generally include national or provincial parks.

Commercial land use ranges from marginal residential conditions (e.g., local gas stations), to those that approach industrial activities (e.g., warehouses). The CCME describes commercial land as including managed areas (e.g., cultivated lawns, flowerbeds) but not natural areas (e.g., forests). The organisms predicted for these lands are similar to those in residential/parkland areas, but land use does not depend on maintaining the same degree of ecological functioning.

For ***industrial land use***, it is once again assumed that activities do not rely on protecting organisms to the same degree as in agricultural and residential/parkland uses. However, the CCME does not recommend that industrial lands should fail to support any ecological activity, nor should they be viewed as a place for high levels of contamination. Soil quality guidelines were developed for industrial land use, but at a lower level of protection.

5.6.5 Development of Environmental Quality Objectives

National guidelines provide broad guidance, but protecting water, sediment or soil quality in a specific site or region might require site-specific *objectives*. Such objectives focus the broad guidelines on the needs of a particular site, and might modify them for local conditions. Economic and social factors might also be incorporated. Guidance on deriving objectives is provided in the water quality guidelines for Canada (CCREM, 1987) and in the USA there are descriptions of policy (USEPA, 1984a) and specific procedures (USEPA, 1994e). Valiela and Whitfield (1988) review the topic briefly.

Local species might be the first factor considered when deriving environmental quality objectives. The USEPA (1994d) provides a fixed procedure for deriving local water quality guidelines by substituting toxicity data for local species in the calculations. Since EPA guidelines are intended to protect 95% of the species, more stringent objectives could be necessary, if it were desired to

Table 11 **Exposure Pathways and Ecological Receptors Considered in Deriving Soil Quality Criteria for the Four Categories of Land Use (CCME, 1996b)**

	Land Use			
Route of Exposure	**Agriculture**	**Residential/ Parkland**	**Commercial**	**Industrial**
Soil contact	Soil nutrient cycling processes	Soil nutrient cycling processes	Soil nutrient cycling processes	Soil nutrient cycling processes
	Soil invertebrates	Soil invertebrates	Soil invertebrates	Soil invertebrates
	Crops/plants	Plants	Plants	Plants
	Livestock/wildlife	Wildlife	Wildlife	Wildlife
Soil and food ingestion	Herbivores			

protect a species that was among the most sensitive 5%. For example, the USEPA method was used to derive a site-specific objective for ammonia in a Colorado river, specifically allowing for the high sensitivity of the johnny darter, a small fish previously untested for ammonia tolerance (Nimmo *et al.*, 1989).

Physicochemical characteristics of the water might affect the biological availability or toxicity of the material of interest. An example would be decreased metal toxicity in water with if the metals were bound by a high organic content. (Increased metal toxicity could occur if the local water was very soft, but that effect is taken into account in most modern criteria.)

Seasonal changes could also be a consideration in designing quality objectives. There might be fluctuation in variables that modified toxic action, such as the metal-binding organic matter. There is also the seasonal presence of key species and life stages. It is recommended that seasonal adjustment of objectives should seldom be done, and then only with great caution. Any adjustment should lean towards increased stringency during seasons known as times of migration, reproduction, or sensitive life stages. It would be ecologically dangerous to relax a well-documented objective for a reason such as the end of the breeding season of a prominent fish.

One example of careful seasonal modification of objectives was the creation of ammonia objectives for the Flathead River in British Columbia, which adjusted the guideline of CCREM (1987) for total ammonia (MacDonald *et al.*, 1987). The objective accounted for (1) the unusual sensitivity of mountain whitefish to ammonia, (2) life-cycle differences in sensitivity, and protection of spawning and rearing seasons, and (3) seasonal differences in discharge, temperature and pH, all affecting the toxicity of ammonia or exposure of the fish. The resulting objective was specific for each month and was fine-tuned to ensure adequate protection of the whitefish, and by extension all other less sensitive species.

The **background level** of the toxicant might be another factor to consider, but it would not elevate an objective. Background levels of contamination produced by historical human interventions should not be used as a basis for modifying guidelines. It is also a policy of the CCME that water, sediment, or soil quality that is better than Canadian guidelines should not be permitted to decline (CCME, 1991b, 1993a). Although communities exposed to low levels of certain substances tend to adapt to them metabolically, and the organisms might appear more tolerant, not all exposures to toxicants produce adaptation, and allowance should not be made for the phenomenon (see Section 4.15.2).

Cumulative effects of low concentrations of some materials might only be evident after long exposure [e.g., lesions in bottom-dwelling fish exposed to PAHs in sediments (Valiela and Whitfield, 1988)]. For such substances, environmental quality objectives should specify average values to be met over long periods, in addition to instantaneous maxima. National guidelines sometimes specify such chronic limits for substances.

Other sources of stress should always be considered. These might include other toxic chemicals, or non-chemical sources like extremes of weather, disturbance, harvesting, or disease. Multiple toxicants are the rule rather than the exception in soils and sediments, which tend to accumulate hydrophobic and particle-bound chemicals. At sites where the spectrum of toxicants is known, an extra allowance could be built into an objective to allow for possible additive doses. Populations that are in decline or displaying signs of stress from human interference or other stresses (predation, disease, parasitism, competition) might also warrant a margin of safety beyond the guideline.

5.7 *Assessment and Regulation of New Substances*

Key Guidance

- *"New substances" are subject to regulatory assessment so that the risk of environmental effects can be appropriately managed before the substances are released for use.*
- *One to three acute toxicity tests with aquatic organisms are required by regulation in Canada, as well as physicochemical information that allows prediction of toxicity, persistence, and potential bioaccumulation.*
- *Quality and usefulness of toxicity data can be assessed on the basis of good laboratory practices and adherence to standard methods, toxicity curves, and comparison of results from direct toxicity testing with predictions derived from structure-activity relationships.*
- *Criteria are available for deciding whether further data are needed for evaluating the environmental aspects of a new chemical, but professional judgement also plays a strong role.*
- *A discrete environmental risk assessment is part of an evaluation. Toxicity information is used to predict a no-sublethal-effect concentration in the environment (the* ***concern concentration****). The* ***predicted environmental concentration*** *is then estimated and the two are compared for margin of safety.*
- *An acute-to-chronic ratio of 10 is often used to extrapolate from the laboratory data on acute lethality, to a concern concentration. Environmental conditions can sometimes modify the real-world toxicity.*

New substances are subject to regulatory assessment for ecological and human health hazards before commercialization in Canada and other OECD countries (e.g., U.S., members of the EEC, Japan, and Australia). The objective is to identify problematic substances before they are introduced to the marketplace and subsequently into the environment, so that the risk of environmental effects can be appropriately managed beforehand.

Substances are considered "new" to Canada if they *do not* appear on an inventory of chemicals known as the Domestic Substances List. New substances are regulated under the Canadian Environmental Protection Act (CEPA). An instructive example of Canadian regulation of new polymers, one category of new substances, is provided by Matheson and Atkinson (1997); the approaches in other countries are covered in other chapters of the same book (Hamilton and Sutcliffe, 1997).

Regulations specify information, including test data, that must be provided for any new substance proposed for manufacture in Canada, or for importation. For administrative details and a procedural description of the regulations and their requirements, see the full *Guidelines for the Notification and Testing of New Substances: Chemicals and Polymers, Pursuant to The New Substances Notification Regulations of the Canadian Environmental Protection Act* (EC and HWC, 1993).

5.7.1 *Requirements for Data*

The entire list of environmental variables which might be required for notification is presented in Table 12 along with the desired test methods. Assessing the ecological hazard does not usually include classical mammalian tests such as studies with rodents. Those are reviewed by Health Canada to assess human health hazards. The minimum set of toxicity tests includes acute tests with fish or *Daphnia*, and algae under certain circumstances, although other information can also be considered (Table 12; section a).

Certain physicochemical properties are listed (Table 12; sections c and d), and many can be used to predict the behaviour and fate of the substance in the environment. The notifier must also describe the manufacturing process, production volumes, likely patterns of use, likely pathways into the environment, and other relevant information that exists or is reasonably obtainable (e.g., from the parent company).

5.7.2 Interpreting Data for Environmental Toxicity

In evaluating data on ecotoxicity for notification of a new substance, a regulatory agency must effectively conduct a predictive risk assessment based on limited toxicological information. Evaluations consider not only toxicity, but chemical predictions of fate, persistence, and partitioning processes that predict bioaccumulation.

The *inherent toxicity* of a substance to an organism is effectively a fixed property, because it is a characteristic of the substance. "Inherent" is used in the dictionary sense of innate, or intrinsic. Toxicity is often described in terms of a reciprocal value, i.e., some benchmark toxicity value such as an ICx or a criterion/guideline for environmental quality.

Hazard incorporates exposure in addition to toxicity. In ecotoxicology, a prediction of hazard considers the potential exposure of organisms to chemicals which are at potentially toxic concentrations. A more specific case of predicting hazard would replace *exposure* with *intake* of the toxicant, if appropriate for the substance and situation.

The first step in assessing the hazard of a new chemical is, accordingly, to develop a toxicity profile from which a ***concern concentration*** can be established. The ***predicted environmental concentration*** is then estimated and compared with the concern concentration (Nabholz *et al.*, 1993). This approach is consistent with the principles of the hazard quotient method in ecological risk assessment (Suter, 1993; Calabrese and Baldwin, 1993) and with the process for developing water quality objectives (see Section 5.6.5).

Assessment of new substances might proceed with limited information on toxicity, and wide confidence intervals might exist on the estimate of the concern concentration (Figure 13). During an early stage of the assessment, the predicted environmental concentration might also have wide confidence limits. As more information becomes available and its quality is improved, the confidence intervals would narrow (Figure 13).

When the concern concentration is higher than the predicted environmental concentration, there is a margin of safety, or in terms of ecological risk, a value less than unity for the *hazard quotient* (the ratio of the predicted environmental concentration over the concern concentration). The estimated margin of safety would normally increase as the information base improved, because of the narrowing confidence limits (right side of Figure 13).

If an inadequate set of data did not allow discrimination between the concern concentration and the predicted environmental concentration, a situation of non-confidence would exist in the assessment (left side of Figure 13).

It is important to consider the situation represented in Figure 13, when interpreting toxicity data for new chemicals. The challenge is in understanding where the estimates lie with respect to their true values. The true values can be estimated with increasing precision, but can never be determined exactly. Some approaches for checking the plausibility of the toxicity profile and the concern concentration are covered in the following section.

5.7.3 Establishing the Reliability of the Toxicity Profile

A Canadian notification scheme might provide as many as three tests of environmental toxicity for the toxicity profile, although additional predictions might be available from QSAR. An assessor would first scrutinize the quality of reported data to determine if they were valid. If so, the assessor could proceed to the objective of extrapolating from the laboratory tests to a concentration of concern in the environment. Four components in scrutinizing quality are discussed in the following text.

1. Adherence to methods of the toxicity test. Data submitted under the New Substance Notification Regulations of CEPA must comply with the *OECD Principles of Good Laboratory Practices (GLP)* and should be derived from tests by the methods of OECD, or other acceptable methods. Toxicity tests are expected to

Table 12 **Physical, Chemical, and Toxicological Information Required for Notification of New Substances under the Canadian Environmental Protection Act.**

Requirement for Data	Test Method
a. Toxicity tests (for chemicals and polymers)	
Acute lethality to fish	OECD Guideline 203 or Environment Canada method
Acute toxicity to *Daphnia*	OECD Guideline 202 or Environment Canada method
Toxicity to algae (polymers only)	Environment Canada method or USEPA (1994a)
Ready biodegradability	OECD Guideline 301
b. Mammalian toxicological tests (for chemicals and polymers)	
Acute mammalian toxicity	OECD Guideline, 401, 402, 403
Skin irritation	OECD Guideline 404
Skin sensitization	OECD Guideline 406
Repeated-dose toxicity	OECD Guideline 407, 410, 412
Mutagenicity	OECD Guideline 471, 473, 474, 475
c. Physicochemical tests (for chemicals)	
Melting point	OECD Guideline 102
Boiling point	OECD Guideline 103
Density	OECD Guideline 109
Vapour pressure	OECD Guideline 104
Water solubility	OECD Guideline 105
Octanol–water partition coefficient	OECD Guideline 107
IR, UV, Mass, or NMR spectrum	As appropriate
Dissociation constant(s)	OECD Guideline 112
Adsorption-desorption (screening test)	OECD Guideline 106
Hydrolysis as function of pH (preliminary test)	OECD Guideline 111
Particle sizes	OECD Guideline 110
Fat solubility	OECD Guideline 116
d. Physicochemical tests (for polymers)	
Number-average molecular weight Residual constituents with mw <500 daltons and <1000 daltons Water solubility Hydrolysis as a function of pH Ultraviolet spectrum n–Octanol solubility	See Appendix 6 of the Guidelines for Notification (EC and HWC, 1993)

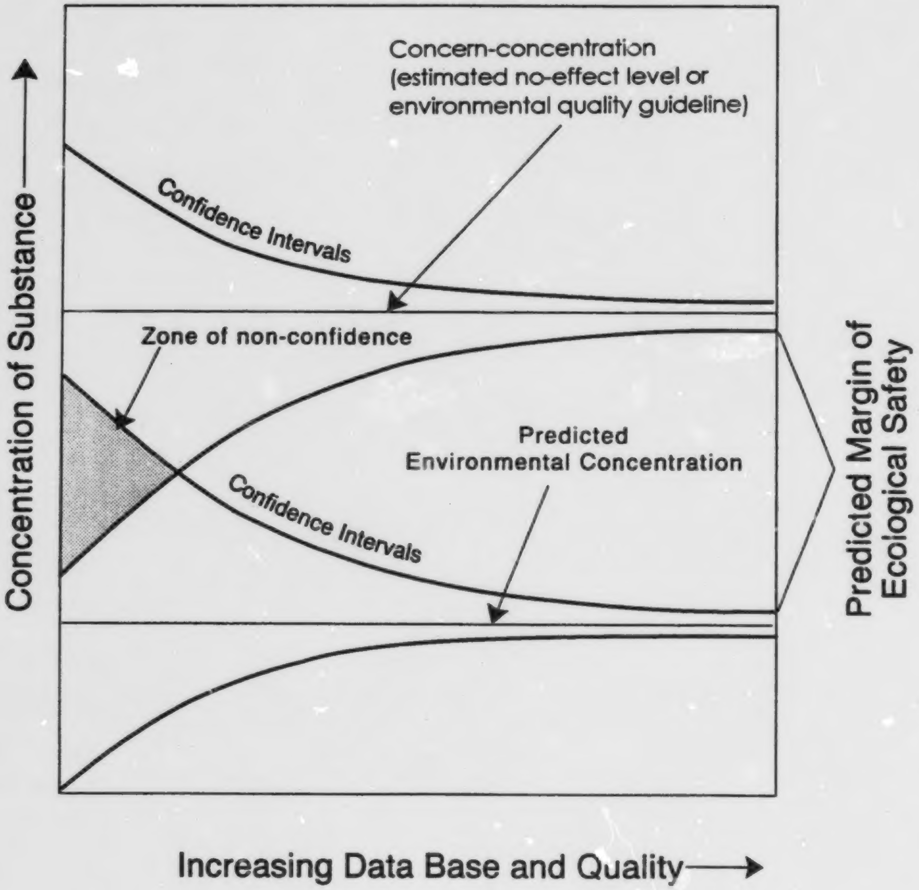
Concern-concentration
(estimated no-effect level or
environmental quality guideline)
Confidence Intervals
Zone of non-confidence
Predicted
Environmental Concentration
Confidence Intervals
Concentration of Substance
Predicted Margin of
Ecological Safety
Increasing Data Base and Quality

Figure 13 **Sequential Risk-assessment Procedure and Predicted Margin of Safety.** This shows the increasingly narrow confidence limits for estimates of the "concern concentration" and the predicted environmental concentration, as the base of information improves in amount and quality. The "concern concentration" might be an experimentally derived no-effect concentration or ICx, or an environmental quality quideline incorporating safety factors. Modified from Cairns *et al.* (1978).

meet GLP, given recent efforts in standardization and programs in laboratory accreditation. If a study did not meet the requirements of GLP, an assessor should review details of how the test was conducted, to judge the reliability of results for assessment purposes.

To establish the reliability of an individual toxicity test, the most relevant performance-based criteria (excluding GLP) are those stipulated within the OECD test methods (e.g., OECD, 1992d) and within the test methods published by Environment Canada (EC, 1990a,c; 1992d), and related guidance documents (EC, 1990d). The criteria can be thought of as falling into seven major areas: test conditions and procedures; water quality; test concentrations; controls and criteria; endpoints; dose-response relationships; and analyses of data (Smrchek *et al.*, 1993; see Section 3).

Professional judgement must be exercised in evaluating test validity. An assessor must decide what frequency can be tolerated for deviations from the standard method, but some violations will carry more weight than others. For example, a single minor excursion outside the desired temperature range might not seriously compromise the quality of a test. Incorrect temperatures throughout the test , however, might alter uptake rate of the toxicant, or accentuate stress, and be unacceptable.

There should be no compromise for some key criteria. Any deviation that significantly alters bioavailability of the test substance should not be accepted. Tests based on nominal concentrations (i.e., unmeasured) should not normally be accepted, since the true exposures could be seriously misrepresented, leading to a misleading endpoint. Deviations from nominal concentrations might occur through human error, equipment failure or physicochemical characteristics. For a mixture, a selected constituent or constituents should have been measured to verify dilutions. For test substances that are volatile, or transformed through hydrolysis or other mechanisms, measured concentrations are imperative. In the presence of extraneous binding agents such as dissolved organic carbon or iron, the chemical method must be designed to measure the dissolved or bioavailable form of the toxicant.

2. Time-related patterns of toxicity. In an acute lethality test with aquatic organisms, the most meaningful endpoint is the *incipient* LC50, also called the *time-independent* or *threshold* lethal concentration. This LC50 would not change its value with longer exposure time, because acute lethal action has essentially ceased (Fry, 1947). Whether the incipient LC50 has been attained is best determined by inspecting a toxicity curve (Figure 14). Although these are seldom constructed in current practice, it would be advantageous to start using them, and such time-series analysis is strongly recommended by an international workshop on statistical analysis of toxicity tests (OECD, 1998a). The 96-h LC50 is usually reported as a convenient and reasonable endpoint, and for many substances it approximates the incipient LC50 (Sprague, 1969). That is not always the case, however, and a longer time might be required to fully reveal the extent of acute toxicity, for substances that were only slowly absorbed (see Section 6.4).

It is important to be aware that acute lethality could be underestimated, even if a test complied with GLP. Results should be scrutinized to see if mortalities continued late into the exposure period. If so, additional mortalities might have occurred in a longer exposure, resulting in a lower LC50 and, probably, a lower concern concentration.

For example, in Figure 14, Toxicant A appears to have reached an asymptote with the time axis within 96 hours. Mortality had effectively ceased and the 96-h LC50 is a useful endpoint, an estimate of the incipient LC50. Toxicant B does not show an asymptote, and continued acute mortality would have been likely with longer exposures. The 96-h LC50 would be an underestimate of acute lethality. Such differences in response to toxicants are clearly important for establishing the toxicity profile for assessing new substances. It would be desirable to design acute tests so that toxicity curves could be depicted and interpreted.

3. Reliability of estimates from QSAR. QSARs (see Section 4.6.1) might provide much-needed information for assessing new substances. Within CEPA, estimates from QSAR may be submitted to augment the results of toxicity tests or possibly in lieu of such tests. Thus, it is important to establish some measure of confidence that the predicted toxicity is valid. In the USA, QSARs have been heavily used in new substance assessment under the Toxic Substances Control Act, which historically has not required preliminary toxicity testing of new substances.

The OECD (1992a) has recommended criteria for judging reliability of using QSARs to estimate the probable toxicity of a new substance (Section 4.6). The OECD considered it most important to establish whether the QSAR being applied to the chemical had originally been developed from chemicals with similar structure and mode of action.

An estimate of confidence in the predicted toxicity should also be reported (e.g., the coefficient of determination (r^2) for the original QSAR). Various computer programs can generate toxicity estimates from QSARs, and they usually provide a description of confidence in the estimate (e.g., USEPA, 1994c).

4. QSAR and measured toxicity as mutual checks. QSARs can provide an indirect means of evaluating results from toxicity tests. For example, the assessor could determine whether a measured LC50 agreed with a QSAR or was an outlier. Concordance would strengthen the apparent reliability of each, but an assessor should be cautious about accepting a measured LC50 that is an outlier of a QSAR. Inconsistencies of measured LC50s

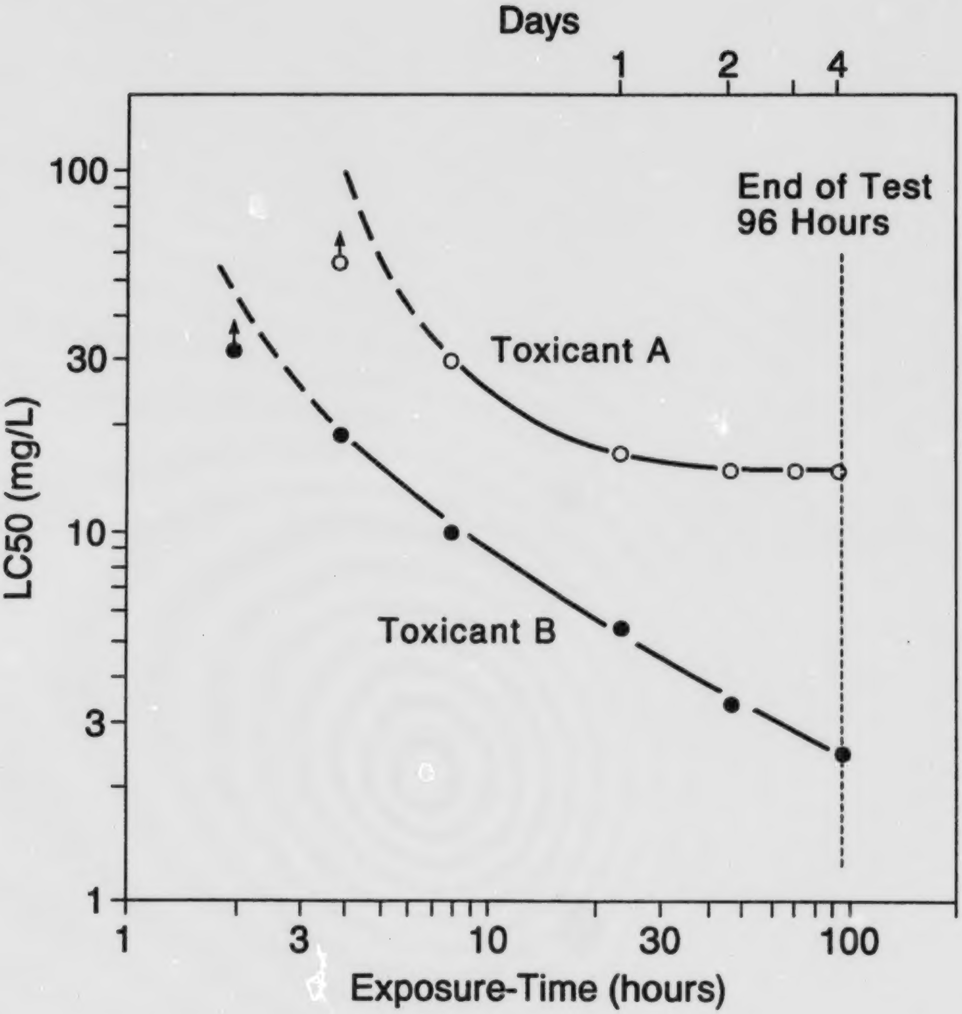

Figure 14 **Toxicity Curves for Two Hypothetical Toxicants.** Each curve was obtained by eye-fitting a line to a series of LC50s, which were derived from inspections of mortality at successive exposure times (from Sprague, 1985).

might arise for reasons described in the preceding three items (e.g., overly short exposure times, or use of nominal concentrations).

If the measured toxicity and the QSAR estimate are both considered valid, yet are quite different, it suggests a unique mode of action for the substance, different from the basis of the QSAR [54]. In case of a clear difference, the estimate by QSAR should be regarded with suspicion since it would conflict with an apparently valid LC50.

5.7.4 *Determining the Need for Further Studies*

Professional judgement again comes into play in deciding whether the level of environmental concern can be estimated with a reasonable degree of confidence. If the available toxicity data are considered valid, the next question is whether they are sufficient for deriving an environmental concern level, or whether additional studies would be desirable. Judgement is required since a very toxic substance would deserve more information than a mildly toxic one.

---

[54] An easily applied initial criterion of difference could be absence of an overlap in the 95% confidence limits of the measured LC50 and those of the toxic concentration estimated by QSAR.

To help standardize this process, decision criteria have been provided by U.S., German, and international regulatory agencies (Walker, 1990; Greim *et al.*, 1993; OECD, 1993a). All these methods require that decisions should be based not only on toxicity, but also on persistence and bioaccumulative potential. That is consistent with the principles of ecological hazard assessment under CEPA. Thus, substances with marginal data, which are suspected of being highly toxic, highly persistent, and capable of significant bioaccumulation, would readily qualify for additional investigation. Different agencies have used various specific criteria for these characteristics, to decide when further toxicity information is needed. Usually, the initial requirement is for information on acute toxicity, which is comparable to that required under CEPA.

5.7.5 Extrapolation to Concentrations of Concern

If a profile of laboratory toxicity data is accepted for a new chemical, an assessor must then extrapolate to obtain a concentration of environmental concern. The procedure will be limited to using an acute-to-chronic ratio (ACR, or its inverse, the application factor, see Sections 4.2.1 and 4.2.2) since they are applicable to small data-bases. Distributional methods would require larger data-sets (Section 4.2.4).

The typical approach divides the lethal concentration by a default value of 10, to obtain an estimated sublethal concentration of concern (Nabholz *et al.*, 1993). That procedure has been validated to some extent by comparing results from distributional methods (OECD, 1992b). Calabrese and Baldwin (1993) suggested that for substances showing slow uptake, a steady internal state might not be achieved during a test of acute lethality, and a divisor >10 might be more appropriate as an ACR. If there were different modes of action for lethal and sublethal effects, extrapolation using an arbitrary ACR might not be suitable. Anilines, for example, are more appropriately assigned a larger ACR (OECD, 1992b), which is consistent with their lack of concordance with QSAR (OECD, 1992a).

Figure 15 helps to visualize how extrapolations from the laboratory to the field might yield either an overly conservative margin of safety, or one that was not large enough for protection. The left side of the figure presents an hypothetical situation in which the concern concentration (derived from effects in the laboratory) is relatively close to the predicted environmental concentration. However, other field conditions mitigate toxicity by chelating the substance and reducing bioavailability. The result is an increase in the true margin of safety. The right side of Figure 15 illustrates the opposite situation. A predicted large margin of safety is reduced in the field by factors which exacerbate toxicity. The example in the figure involves two or more toxicants or agents of stress working together (*synergism*, not necessarily implying a more-than-additive toxicity). The condition could also occur through modifying factors such as soft water which would increase the toxicity of metals.

In summary, a substantial body of empirical evidence indicates that an extrapolation factor of 10 applied to tests of acute lethality on aquatic organisms, has generally provided a reliable and moderately conservative estimate of thresholds of sublethal effect in the environment (Sections 4.2.1, 4.2.2, and 4.1). However, there might be factors that would increase or decrease toxicity, making a larger or smaller ACR appropriate for evaluating new substances.

5.8 Assessment of Priority Substances under CEPA

Key Guidance

- *Substances on the Priority Substances List must be assessed to determine if they are "toxic" as defined in Section 11 of the Canadian Environmental Protection Act.*

- *Environment Canada carries out environmental assessments, while Health Canada does human health risk assessments.*

- *Environmental assessments are carried out in three major steps: problem formulation; analysis of entry, exposure, and effects; and risk characterization.*

- *Effect characterization determines a critical toxicity value for each kind of organism (= assessment endpoint) considered to be at risk. These values are used in risk characterization to calculate estimated no-effect values, which in turn, are compared with estimated exposures to determine whether a substance is "toxic" under CEPA.*

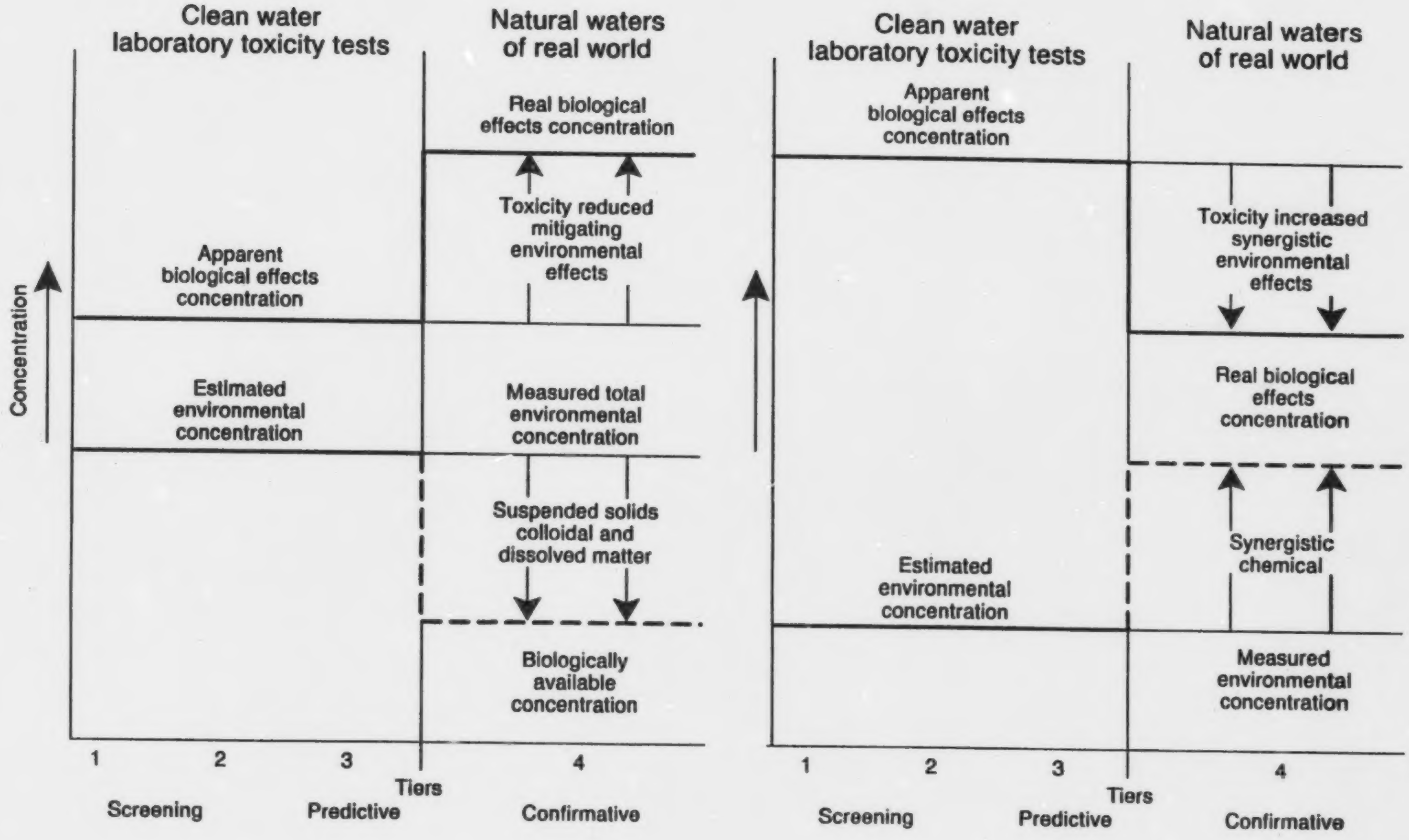
Clean water laboratory toxicity tests
Natural waters of real world
Real biological effects concentration
Toxicity reduced mitigating environmental effects
Apparent biological effects concentration
Concentration
Estimated environmental concentration
Measured total environmental concentration
Suspended solids colloidal and dissolved matter
Biologically available concentration
1
2
3
4
Tiers
Screening
Predictive
Confirmative
Clean water laboratory toxicity tests
Natural waters of real world
Apparent biological effects concentration
Toxicity increased synergistic environmental effects
Real biological effects concentration
Synergistic chemical
Estimated environmental concentration
Measured environmental concentration
1
2
3
4
Tiers
Screening
Predictive
Confirmative

Figure 15 **Hypothetical Situations in which Predicted Margins of Safety are Modified by Conditions Existing in the Field.** On the left, a relatively small margin of safety predicted from laboratory toxicity is actually greater because of factors in the surface water which mitigate toxicity. On the right, a relatively large predicted margin of safety is actually smaller because other toxicants and environmental stressors act in concert with the pollutant of concern. Modified from Cairns *et al.* (1978).

- *Effect characterization can use a weight-of-evidence approach that incorporates some or all of: single- and multispecies toxicity testing; field tests; equilibrium partitioning; and quantitative structure-activity relationships.*
- *Effect characterization can consider diverse biota, including pelagic, benthic, groundwater, and terrestrial organisms, depending on the environmental partitioning of the substance.*
- *Areas of uncertainty must be identified for use in risk characterization.*

5.8.1 Background

Section 12 of the *Canadian Environmental Protection Act (CEPA)* requires the Ministers of Environment and Health to compile and publish a list of substances known as the *Priority Substances List*. Substances on this list must undergo assessments to determine whether they are "toxic" according to the following definition in Section 11 of CEPA.

For the purposes of this Part, a substance is toxic if it is entering or may enter the environment in a quantity or concentration or under conditions

(a) having or that may have an immediate or long-term harmful effect on the environment;

(b) constituting or that may constitute a danger to the environment on which human life depends; or

(c) constituting or that may constitute a danger in Canada to human life or health.

Substances determined to be "toxic" enter a risk management phase in which the federal government, with input from the provinces, industry, and public, determines what measures, if any, will be needed to mitigate effects.

5.8.2 Environmental Assessment

Environmental assessments can be complex. They are concerned with protecting numerous species that might be affected either directly by a substance or indirectly as a result of disruptions to ecosystem structure or function. There is a guidance manual for performing environmental assessments under the priority substances program (EC, 1997f). Sections 5.4 and 5.5.1 of this document discuss environmental assessment and ecological risk assessment.

Environmental assessment of priority substances involves three major steps: problem formulation; analysis of entry, exposure, and effects; and risk characterization. A tiered approach helps to ensure that assessments proceed only to the level of refinement required for effective decisions. Tiers 1 and 2 use, respectively, very conservative and conservative estimates of exposure and effects to determine whether or not a substance is likely to cause harm in the environment. Tier 3 compares exposure and effects distributions, rather than point estimates. The tiered approach is discussed more fully under the heading "risk characterization". A special analysis is used for naturally occurring substances.

Problem formulation uses analysis of pathways and identifies sensitive organisms *(receptors)* that are at high risk of exposure – *assessment endpoints*. Since toxicity data are not always available for assessment endpoints, *measurement endpoints* are used to estimate effects on assessment endpoints. A measurement endpoint is usually obtained by carrying out toxicity tests in the laboratory with related species or by conducting field studies. A *conceptual model* describes the ways in which the substance behaves in the ecosystem and its possible effects. Data gaps that must be filled in order to complete the environmental assessment are identified during problem formulation.

The **analysis** phase consists of three parts: entry, exposure, and effects characterization. The objective of entry characterization is to determine the natural and anthropogenic sources, locations, and quantities of the substance entering the Canadian environment. Entry characterization considers all phases of the substance's life cycle. Information gathered from the characterization of entry may be used to further refine the problem formulation, as input to the characterization of exposure, and in the development of mitigation measures during risk management. The objective of exposure characterization is to determine the *estimated exposure value (EEV)* or the exposure distribution for each assessment endpoint. For wildlife, exposure may be expressed as a total daily intake, as an environmental concentration, or as a tissue residue. For naturally occurring substances, it is necessary to estimate the natural background concentration in areas of concern. The objective of effects characterization is to determine the *critical toxicity value*

(CTV) or the effects distribution for each assessment endpoint. A CTV is a quantitative expression of low toxic effect, such as an IC25. Studies on single- or multispecies toxicity tests, field studies, or critical body burdens can all be used to determine effects. Potential effects of priority substances on stratospheric ozone depletion, ground level ozone formation, and global warming are considered during the effects characterization phase, but these effects will not be further discussed here because they do not directly involve environmental toxicology.

Risk characterization, the third step in environmental assessment of a priority substance, uses a tiered approach. Tier 1 involves dividing the EEV for a worst-case situation by the *estimated no-effects value (ENEV)*. The EEV for the worst-case situation is generally the maximum observed or estimated concentration in the Canadian environment, while the ENEV is calculated by dividing the CTV by a factor to derive a value with a very low probability of causing adverse effects to the assessment endpoint. If EEV/ENEV is <1, the substance is considered not to be "toxic" as defined in Section 11 of CEPA. If the quotient is ≥ 1, the assessment moves to Tier 2. Tier 2 involves a further analysis of exposure and effects to calculate a quotient that is still conservative but more "realistic" than the very conservative quotient calculated in Tier 1. If the Tier 2 quotient is <1, the substance is considered not to be "toxic". If the quotient is ≥ 1, the assessment moves to Tier 3. Tier 3 involves comparison of exposure and effects distributions to obtain a more complete picture of the nature of possible effects in the environment. The substance is considered to be "toxic" if the Tier 3 assessment determines that adverse effects on an assessment endpoint are likely to occur. A special analysis is used for naturally occurring substances. This analysis takes into account natural background concentrations and the possibility of increased tolerance of the substance by organisms normally found in naturally enriched areas.

5.8.3 General Recommendations for Characterization of Effects

Toxicity tests should use Canadian or closely related species in order to minimize uncertainty due to differences in sensitivity of species. Similarly, tests should be carried out under conditions that are pertinent to the Canadian environment. There should be toxicity data from a wide range of trophic levels to help identify the types and magnitude of toxic effects, and the populations, communities, and ecosystem processes that might be particularly sensitive to the substance. When there is a likelihood of chronic exposure to a substance, results from full life-cycle studies are preferred for the derivation of the CTV. In their absence, results from partial life-cycle studies on the most sensitive stages may be used (OECD, 1993d). If only acute toxicity data are available, they may be used as a basis for the CTV (e.g., an LC50, EC50, or acute sublethal IC25).

Toxicity data must be critically evaluated, as only studies of acceptable quality can be used to derive a CTV. (Section 3.1 discusses quality assurance). Whenever possible, toxicity studies using standardized test methods developed by agencies such as EC, the USEPA, and the OECD should be used for effects characterization. If other test methods are used, the procedures must be described in sufficient detail so that the reliability of the results can be evaluated. If no acceptable studies are available, research might be carried out to supply essential toxicological data.

Sources of uncertainty with the toxicological data must be documented. These uncertainties are taken into account in selecting factors or in conducting uncertainty analysis during risk characterization.

5.8.4 Types of Data Used in Characterization of Effects

CTVs are most commonly derived from the results of single-species toxicity testing. Valuable information for characterization of effects can also be provided by field studies, multispecies tests (including microcosms and mesocosms), determinations of critical body burdens, quantitative structure activity relationships (QSARs), and estimates of equilibrium partitioning. When possible, several of these methods should be used together, to estimate risk by a weight-of-evidence approach (see Section 4.4).

Single-species toxicity tests are excellent for determining the toxicity of a substance in a particular medium (e.g., water, sediment, or soil) under specific test conditions. However, single-species toxicity tests are not designed to integrate the simultaneous study of toxicity and various chemical transformation and partitioning processes. The usefulness of toxicity tests for identifying potential effects in the natural environment depends on the degree to which results can be extrapolated directly to natural systems and also on how well the tests can be replicated (Cairns, 1992).

Field studies and multispecies tests using microcosms or mesocosms can be used to determine toxic effects under conditions that more closely mimic the natural environment. They can also reveal secondary effects that result from species interactions (OECD, 1995a; Sections 4.1 and 6.3 herein). Field studies are particularly useful in the environmental assessment of mixtures and effluents, where direct cause-and-effect relationships might be difficult to ascertain in the laboratory. However, field and multispecies studies have some drawbacks. They require a period of time for component species to stabilize, and they are usually more costly than single-species toxicity tests (USEPA, 1992a). Natural communities are often difficult to sustain, and extinctions and changes in community structure might occur independently of exposure to the test toxicant (Buikema and Voshell, 1993). It might be difficult to determine the effects of a toxicant in field studies because of differences in habitat quality between areas, natural variations in environmental conditions within areas, the occurrence of undetected stressors, and the movement of organisms into or out of the study area (USEPA, 1992b).

A *critical body burden (CBB)* is the minimum tissue concentration associated with an adverse effect in an organism. CBB can be a useful endpoint in acute toxicity studies, since there is a direct measurement of body burden, without the problems associated with estimating bioavailability and accumulation. When available, CBBs should be compared to data on tissue residues or body burdens collected in the field. Such data may be used as the basis for the risk assessment or to support other apparent effects, in a weight-of-evidence approach. While CBBs of organic substances have been linked to acute toxicity of narcotic chemicals in aquatic organisms, more research is required before this concept can be generally applied to other modes of toxic action (McCarty and Mackay, 1993; Section 6.4 herein).

Quantitative structure activity relationships (QSARs; Section 4.6) may be used to predict effects of chemical substances in the absence of empirical data. They are developed for groups of substances that are differentiated by mode of action, which varies with the structure and physical-chemical properties of the substances. QSARs are only applicable to substances within their particular group. QSARs can be used to make preliminary estimates of toxicity in problem formulation, to corroborate empirical data, to determine the need for additional testing, or as part of a weight-of-evidence approach for selecting a particular CTV.

The *Equilibrium Partitioning (EqP)* method estimates the toxicity of hydrophobic, nonpolar, non-ionic organic substances to benthic or soil-dwelling organisms using toxicity data for free-swimming organisms (Di Toro *et al.*, 1991; van de Plassche and Bockting, 1993). The K_{ow} of the substance and the organic carbon content of the solid medium must be known, and it is assumed that a chemical equilibrium exists among porewater, organic carbon in solid phases, and resident biota. It is also assumed that organisms in the contaminated medium are equally sensitive as the free-swimming organisms, to the substance. The uncertainty of these assumptions can often limit the method's usefulness (Chapman, 1989b). Effect values derived by EqP may be used as screening values during problem formulation, or as part of a weight-of-evidence approach for selecting a particular CTV.

5.8.5 Deriving Critical Toxicity Values

In order of preference, the CTV may be in the form of an IC25 (or lower value such as IC10 if reliably derived); a Lowest-Observed-Effect Level (LOEL) if an IC25 cannot be calculated; or a median effect value (e.g., LC50) if an IC25 or LOEL cannot be derived.

It is advantageous to estimate CTVs by means of point estimates (e.g., IC25) rather than by estimates such as NOEC/LOEC derived from testing hypotheses. The advantages are described in Sections 2.5.2 and 2.5.3, and have been pointed out by many authors (e.g., Stephan and Rogers, 1985; Pack, 1993; Suter, 1996).

A major disadvantage of endpoints obtained by testing hypotheses is that the LOEL is limited to one of the concentrations that happened to be chosen for testing. An LOEC would not necessarily correspond to any consistent effect-level from one test to the next, but might, for example, represent 5% effect in one test and >50% effect in another. Poor experimental design or procedure might produce an erroneously low estimate of toxicity for a substance, since even large decreases in survival or growth might not be statistically significant. Valuable information from the dose-response curve, such as slope, is not used in testing hypotheses. As a final disadvantage, no confidence limits can be estimated for the LOEL.

By contrast, point estimates can be estimated as any concentration within the range tested. Poor experimental design or procedure will be reflected in wider confidence limits, but will not affect the magnitude of the estimated ICp. Point estimates require that a dose-response model should be specified, whether logistic, probit, or multistage. Superior approaches would use that model with regression analysis to obtain the desired point estimate of effect.

The following points provide summary guidance on how to derive a CTV (see EC, 1997f for more detail). For any given approach or method, assessors must ensure that assumptions have been met and limitations understood.

- Dose-response curves with confidence limits should be estimated for toxicity studies used to derive the CTV.
- The *p* in the ICp from the estimated dose-response curve should be no lower than 25, unless it is being estimated by interpolation (Moore and Caux, 1997).
- Generally, a non-linear regression analysis is preferred over a weighted linear regression analysis. Concentrations or doses should be transformed as $\log_{10}$ unless there is a compelling reason not to do so. Model adequacy must be tested with a goodness-of-fit statistic and, if the model fit is inadequate ($p < 0.05$), the results should not be used.
- If a LOEL or NOEL is used as the CTV, information about the number of replicates, test variance, α, β, and test dose intervals should be provided. Generally, LOELs are preferred to NOELs, and TOECs are not used to derive a CTV.

5.8.6 Types of Biota and Environmental Compartments of Concern

For evaluating a substance in aquatic habitats, most data on pelagic biota will probably result from single-species toxicity tests. For these organisms in particular, a large number of standardized test methods have been developed (Section 3.2.1). Results from multispecies tests and field studies can also be useful for characterizing direct and indirect effects of substances on pelagic biota under natural or near-natural conditions. An approach using critical body burdens (CBB) is particularly relevant when it is difficult to determine the concentration of bioavailable forms of a substance in the aquatic environment.

For freshwater, marine, and estuarine sediments, a number of approaches have been developed to evaluate the toxicological significance of substances (VKI and RIVM, 1995; Section 4.11). Spiked-sediment toxicity tests establish cause-and-effect relationships between exposed organisms and spiked concentrations of individual substances or mixtures. When no suitable spiked sediment toxicity test results are available, a weight-of-evidence approach should be used to establish associations between a substance's concentrations in sediments and observed adverse biological effects. These associations can be based on co-occurrence data from laboratory tests conducted on field-collected sediments that contain mixtures of substances. The CCME (1995) provides a discussion of the co-occurrence approach based on work by Long (1992), Long and Morgan (1990), and Long and MacDonald (1992). Other types of co-occurrence approaches include the *apparent effects threshold (AET)*, sediment quality triad, and informal evaluations of chemistry and biological responses (USEPA, 1992c). The benthic community structure assessment is another approach that may be used to compare a community living at a reference station with a community living in a contaminated area. (Examples of this approach are given by Diaz, 1992; La Point and Fairchild, 1992; Persaud *et al.*, 1992; Reynoldson and Zarull, 1993; Reynoldson *et al.*, 1995.) The EqP approach (see Section 5.8.4) may be used when the sediment solid phase contains more than 0.2 % organic carbon.

Groundwater can contain a biologically diverse community, including bacteria, protozoa, and micro- and macro-invertebrates, especially copepods and amphipods (Botosaneanu, 1986; Danielopol, 1992; Marmonier *et al.*, 1993). No standard methods for toxicity tests exist for groundwater organisms, and the literature is limited to descriptions of effects on bacterial mineralization and acute toxicity tests with groundwater invertebrates (Notenboom *et al.*, 1994). When groundwater toxicity data are unavailable, surrogate species such as surface water crustaceans may be used to estimate a CTV for functionally similar species (Notenboom *et al.*, 1994; Section 4.12.3).

Terrestrial (soil) biota include, for assessment purposes, plants, microorganisms, invertebrates that live at least part of their life cycle in the soil. Mammals, birds, reptiles, and amphibians are assessed separately as wildlife (see next paragraph). There are various approaches for assessing the effects of priority substances on terrestrial biota, including single-species and multispecies toxicity tests, and field studies. Methods have been developed to assess toxic effects on earthworms and terrestrial plants (OECD, 1993d; USEPA, 1985b; see also Section 4.12). The International Organization for Biological Control has many standard methods for assessing effects on terrestrial insects. Terrestrial organisms can acquire toxic substances by ingestion, dermal uptake, or inhalation of gases and fine particulate matter, and plants can also take in substances through the roots or leaves. Therefore, the partitioning of the substance within terrestrial compartments, and the life habits of the biota must be considered, to determine the relevance of results from toxicity tests. If there is a lack of toxicity data for soil organisms that are exposed to a substance primarily via soil porewater, then toxicity data for aquatic species may be used to estimate effects (VKI, 1994).

For assessment purposes, "wildlife" refers to wild mammals, birds, amphibians, and reptiles. Environment Canada has developed a model to estimate a substance's daily intake rates for wildlife, under conditions of multimedium exposure. Field studies are preferred for deriving a CTV when cause-and-effect relationships can be clearly established. When field studies are unavailable, laboratory studies may be used, preferably ones which actually use a wildlife species. Methods for testing wildlife have been reviewed by Hoffman *et al.* (1995). Mammalian wildlife assessments rely heavily on laboratory data generated for human assessments (Hodgson and Levi, 1987), although the USEPA has test methods for mink (*Mustela vison*) and European ferret (*Mustela putorius furo*). Methods for birds include tests that are acute, short-term dietary, chronic reproductive, embryotoxic/teratogenic, behavioural, and field-oriented. There are no standard methods for amphibians.

For wildlife, data on reproductive and developmental toxicity and reduced survival are preferred since they can be directly related to potential effects on populations. Biochemical or physiological perturbations such as endocrine disruption, genotoxicity, and immune suppression might also have serious repercussions for wildlife populations. The Canadian Wildlife Service of Environment Canada has established a working group to address these issues. In addition to considering direct toxic effects, environmental assessments must also take into account the indirect effects a substance might have on wildlife through behavioral alterations, decreased food supply, or habitat degradation.

5.9 Evaluation of Consumer Products

Key Guidance

- *The Environmental Choice™ Program of Environment Canada is designed to assist consumers in choosing manufactured products having minimum environmental impact. It is also designed to encourage environmentally sound practice within industries.*
- *The rating process for a given category of product includes examining the life cycle of the product: energy and materials used to make it; amount and quality of wastes during manufacture; environmental effects of the product; and impact at the time of disposal.*
- *Toxicity tests are one element of the rating. They can be tests of the product, or of the waste from making it (e.g., next items).*
- *General purpose cleaners are tested for lethal and sublethal toxicity with a battery of aquatic tests. The realistic test concentrations represent normal to extreme use.*
- *Mills producing sanitary paper products a product must meet Canadian effluent regulations. Mills are then rated for various qualities, including sublethal toxicity of the effluent to fish. The result is expressed as the Toxic Emission Factor, i.e., the degree of toxicity multiplied by the volume of discharge per unit of product.*

The goal of the Environmental Choice™ Program (ECP) of Environment Canada is to establish guidelines for consumer products which are environmentally friendly. Products which meet the guidelines are identified through an official EcoLogo label as more environmentally benign than others in their class. The program was designed to allow discriminating consumers, who cannot investigate the environmental effects of every product, to choose products with the least environmental impact.

The program also provides industry with a market incentive to improve manufacturing processes. The ECP considers the complete life cycle of the product: energy and materials used in manufacturing, amount and quality of wastes, environmental effects of the product, and its eventual disposal. Some guidelines set criteria for only one or two categories. In other cases, criteria are set for each of the relevant categories, points are assigned for each, and the total score is called the *load point*. Thus the load point system is the summary index of the major environmental effects associated with a product. The criteria are designed with the intention of including those products that score in the top 20% for environmental performance. Products that meet the criteria are allowed to display on the packaging, the *EcoLogo*™ which is a pattern of three intersecting doves in the form of a maple leaf.

The most comprehensive criteria are those being used for sanitary paper products. The guidelines cover products such as toilet paper, paper towels, facial tissues, and table napkins. The rating system includes information on the toxicity of the effluent and its chemical oxygen demand, energy consumption, natural resources utilization, and solid waste generation.

Results of toxicity testing have been used or considered under the ECP for several classes of products such as general purpose cleaners, paper products, two-stroke marine engine oil, chain-saw lubricants. The following section describes the general use of toxicity in ECP, and give examples for two categories of products.

5.9.1 Toxicity Testing in Product Evaluation

Toxicity data must be carefully used for a fair rating of the products and their manufacturing processes. Generally, two kinds of toxicity testing can be used: tests on the products themselves (e.g., general purpose cleaners) and tests on effluents from the factories (e.g., pulp and paper). Summarizing that toxicity information for ECP raises the same issues as in creating a toxicity index (Section 4.5). The items of concern are whether to consider both lethal and sublethal toxicity, the relative weights to be given to results of different tests, assessing the mass of toxicant discharged, and incorporating the frequency and duration of toxic episodes.

General purpose cleaners. Toxicity tests for rating cleaners follow the basic principles described in Section 2. In particular, standard methods are cited and evaluation relies on a battery of aquatic tests comprising three or more species from different phyla and trophic levels, (e.g., a bacterium or an alga, an invertebrate, and a fish). Lethal and sublethal endpoints are included. While lethality is clearly detrimental, the more likely effect of general purpose cleaners in nature will be long-term sublethal effects on growth, reproductive physiology and performance, or other aspects of health. Some of the standard methods such as the seven-day *Ceriodaphnia* test, allow both lethal and sublethal endpoints to be assessed in one test.

Toxicity testing of general purpose cleaners may be done on the individual ingredients, or on the whole formulation using the amount or dilution recommended by the manufacturer. The toxicity levels for whole-formulation testing are based on data from tests conducted on ten representative products by Centre St. Laurent of Environment Canada (Bermingham *et al.*, 1995).

The concentration tested must be reasonable, since any substance can be toxic if the concentration is great enough. There is little point in comparing products at concentrations that would never occur in the environment. Extremely high concentrations are likely only from spills, which would be severe but local and brief. General purpose cleaners are tested over a range of concentrations that could conceivably arise from normal domestic or institutional use of the product. The range embraces the normal dilution of the product, which can be calculated from instructions for the product, plus an allowance for excessively strong concentrations.

Sanitary paper products. Toxicity of the effluent is one of the criteria for pulp mills, under the ECP for sanitary paper products. The procedure requires that: (1) mills which do not comply with all federal effluent regulations, including absence of acute toxicity, should be automatically disqualified for the ECP; and (2) the eligible mills are to be assessed by sublethal toxicity tests since they would be discharging nonlethal effluent. The sublethal test is with fathead minnows for mills discharging to fresh water, and with inland silversides for mills discharging to marine or estuarine waters. Since

Canadian mills are already required to perform these tests quarterly on their effluent (Section 5.2), no additional tests are imposed.

Other sublethal tests are part of regular monitoring of pulp and paper effluents under federal regulations (Section 5.2), but were not included in the Environmental Choice™ Program. Data for those other tests, from the early cycles of environmental monitoring might indicate benefits from including them in a future version of the sanitary paper products program, or in revisions of the other paper programs.

In this program for sanitary paper products, all toxicity data are weighted equally. Endpoints are converted to Sublethal Toxic Units (Section 4.3.1). The mass of toxic material likely to enter the environment is a key consideration. The program explicitly considers effluent volumes as recommended by McCubbin and Folke (1993). Effluent toxicity is expressed as the standard Toxicity Emission Factor (TEF) discussed in Section 4.3.3. This is based on a sublethal endpoint.

TEF = (100/NOEC) × (effluent flow in m^3 /tonne of product)

Thus, the rating measures the amount of sublethally toxic material discharged per cubic metre of effluent, normalized for the size of the mill. The philosophy is that a mill should not be penalized if it reduces its use of water, leading to a more concentrated effluent. For example, if mill A had an effluent that was twice as toxic as mill B (measured as IC25, i.e., a percentage concentration of effluent), but mill A used only half as much water as mill B, per tonne of product, then the two mills would rank equally for TEF. That is equitable, since they would be discharging the same amounts of toxic substance per unit of production.

Emerging Issues

The science of ecotoxicology continues to advance, and this section briefly describes some topics that are under study at the research level. Results for some topics might, in the future, provide explanations of the mechanisms of toxic effect and the reasons for observed variation; other topics might provide useful new tests for routine use. Some topics represent tools for monitoring or surveying which are outside the scope of this document, but of interest to environmental toxicologists. The examples are mostly for aquatic organisms, particularly fish. Unfortunately, it did not prove feasible to cover developments for other habitats and groups of organisms. Each topic is briefly reviewed, some implications are indicated, and some references are supplied for interested readers.

6.1 Genotoxicity

Key Guidance

- *Two approaches using laboratory tests, and one for field surveys, have been commonly used to assess the potential of materials to cause genetic damage.*

- *Tests with micro-organisms measure the interaction between the test material and the genetic material (DNA or RNA). The* ***Mutatox*** *test and the* ***SOS Chromotest*** *are fast, convenient, and suitable for screening materials for genotoxicity in environmental work. However, the tests do not necessarily predict carcinogenicity in complex organisms.*

- *Tests on teratogenicity can measure the incidence of deformities in newly hatched fish following one or two weeks of exposure starting with fertilization. These whole-organism tests are environmentally relevant because they measure the lethal genetic damage at the organism level.*

- *Incidence of tumours or developmental deformities in bottom-living wild fish represent a useful tool for field surveys of genotoxic effects. Tumours are easily recognized, but a number of biological and environmental factors can confound interpretation of results.*

Genotoxicity is the potential of a material to damage the genetic material within an organism. It can be expressed as lethality, reproductive failure, heritable mutations, teratogenesis (birth or hatching defects), or as carcinogenesis (growth of tumours, or more precisely, *neoplastic lesions*[55]). Some short-term laboratory tests are now available, and are suitable for screening use, if there is concern for the genotoxic potential of a particular material.

Rapid advances in genetic and molecular biology can and should be integrated into assessments of environmental toxicity. The background and basics of environmental genotoxicity are reviewed by Shugart (1995), along with some examples. Other instructive background reviews are provided by participants in a workshop (Anderson *et al.*, 1994) and in a report to Canadian government agencies (Anderson *et al.*, 1998). Those documents discuss research approaches that could be used, how these could document and explain subtle effects, and also how they might be used in a predictive manner for ecological effects. Some specific examples of linking genetic damage with the ecological implications of reproductive success are given by Anderson and Wild (1994). For example, they describe tests with the nematode *Caenorhabditis elegans*, whose DNA complement was completely described by late 1998, and whose 959 cells are understood in both structure and function. Exposure to 45 m*M* concentrations of a certain organic toxicant were required to cause a decrease in viable progeny from these nematodes, and yet concentrations as low as 0.5 m*M* caused inherited genetic change.

---

[55] "Tumour" is a general term that includes many kinds of internal or external growths, some of which would be caused by viral agents, and benign growths which are not on a carcinogenic pathway. A neoplastic lesion is a collection of altered cells which show signs of reverting to an undifferentiated cell type. The cells undergo rapid and invasive growth invasive growth. If caused by a toxic chemical, the growth continues after removal of the toxicant.

Tests for genotoxicity can be very important, particularly at the research level; however, they need not necessarily be part of all programs for assessing or monitoring toxicity. Considerable attention has been focussed on this type of potential toxicity, perhaps as a result of public concern about genotoxic effects on humans from contaminants in food and air. In this respect, a case can be made that there has been an over-reaction by the public. One of the central figures in the field of testing genotoxicity, Dr. Bruce Ames, has spoken and written on relative risks, including the statement that "... dietary residues of synthetic pesticides and other traces of synthetic chemicals and industrial pollutants are not likely to be significant carcinogenic risks ..." (Ames, 1966). Environmental toxicologists should maintain a broad but balanced outlook on relative hazards, likely modes of toxicity, and appropriate tests to be selected for a particular situation.

A discussion follows on three approaches: test measuring damage to DNA/RNA, teratogenicity, and surveys of tumours.

6.1.1 Tests Measuring Damage to DNA/RNA

This approach measures the interaction of a test material with the genetic material (DNA or RNA) in micro-organisms or isolated eucaryotic cells (cells of higher organisms, having a nucleus). Several screening tests have been developed with micro-organisms, of which the **Ames test** is the best known and most widely used (Ames *et al.*, 1975; Maron and Ames, 1983). It uses nonvirulent strains of the bacterium *Salmonella typhimurium*, mutated so that they lack the ability to synthesize the amino acid histidine, and therefore can be cultured only on a medium containing histidine. Mutagenic test materials cause some bacterial cells to revert to a strain that can manufacture histidine, and such a reversion results in a colony of bacteria on histidine-free medium.

The Ames test is simple, inexpensive, and fairly fast (L'Haridon *et al.*, 1993). Positive tests reportedly correlate well with production of neoplastic lesions in laboratory animals, but relative potency of materials does not always correlate well with potency for complex organisms. Relevance to carcinogenesis in complex organisms can be increased by adding a rat liver homogenate to the test system. Many potentially mutagenic compounds, in particular PAHs, are transformed into active mutagens by some liver enzymes. Thus the enhanced Ames test incorporates biotransformations that produce mutagens—a better representation of processes within the organs of complex organisms.

The **Mutatox test** uses a dark mutant of the normally bioluminescent bacterium *Vibrio fischeri*. Exposure to genotoxicants induces mutations, and some bacteria regain the ability to produce light. Measuring light output in a photometer provides an estimate of genotoxicity. The measuring system is similar to that used in the toxicity test with luminescent bacteria, and is offered by the same company (Microbics, 1993).

The **SOS-Chromotest** is available commercially. In bacteria, SOS signifies an inducible enzyme system that repairs lesions in bacterial DNA (Walker, 1984). The Chromotest uses a genetically engineered strain of *Escherichia coli* in which the enzyme ß-galactosidase has been linked to the genes producing SOS proteins, although the enzyme is normally unrelated to SOS repair mechanisms. Hence, if a bacterium suffers genetic damage that induces the SOS repair mechanism, ß-galactosidase will also be produced and can be measured colorimetrically (Quillardet *et al.*, 1982; 1985). A parallel measurement of cellular metabolism is based on the activity of alkaline phosphatase, also measured colorimetrically; it estimates cell mortality, so that can be separated from any genotoxic effects. A given test thus provides two distinct endpoints, one which measures the genotoxic potential of the test material (production of ß-galactosidase), and another which measures the toxicity (activity of alkaline phosphatase).

The SOS-Chromotest and Mutatox tests are both convenient as screening tests for genotoxicity, in environmental studies. Both are simple, require only a few hours and small samples, and can be easily quantified. Large numbers of samples can be processed in the SOS-Chromotest by using 72-well microplates. Both tests can be run with added homogenate of rat liver, to detect indirect mutagens. Both get a favourable rating for environmental studies compared to the more lengthy Ames test, in a comparative study by Legault *et al.* (1994).

Some users have found it difficult to accurately judge the colours produced, and measurement with a colorimetric plate reader would be appropriate. The SOS-Chromotest was used to investigate genotoxic agents in water and sediments of the St. Lawrence river (Langevin *et al.*, 1992). The use of the SOS Chromotest is reviewed in a thorough fashion by White and Côté (1998) and they conclude that the semi-automated version of the test can efficiently examine samples required for general conclusions about the sources and ecological behaviour of genotoxic organic contaminants in aquatic systems.

All three micro-tests have some limitations. Concentration of environmental samples is often necessary because genotoxic agents are often in low concentrations. None of them can predict carcinogenicity in complex organisms, instead they are intended as screening assays to detect compounds with genotoxic potential. Because complex organisms have many layers of defense to prevent mutations from being expressed, a positive response in the SOS-Chromotest (production of ß-galactosidase) indicates only genetic damage, which might not have led to a mutation. Positive responses in the Ames and Mutatox tests indicate that a heritable mutation has occurred.

Several screening tests for environmental mutagens are based on genetic aberrations in isolated cells of various multi-cellular organisms. The **Sister Chromatid Exchange Test** can score the frequency of different chromosomal aberrations as a function of the concentration of test material. A similar test looks for aberrations during division of fish cells (Kocan *et al.*, 1982). Many biochemical probes for DNA alterations have been developed and are reviewed in Shugart *et al.* (1992).

Some short-term tests of genotoxicity use multicellular organisms such as fish embryos (Herbert and Hansen, 1998). Some classes of genotoxic agents disrupt DNA replication during cell division (mitosis) of fish and amphibians, and lead to the formation of micronuclei in red blood cells. For example, one test procedure exposes larval newts for 12 d, to renewed test solutions in flasks. The incidence of micronuclei per 1000 red blood cells is counted in blood smears from treated and control animals (Jaylet *et al.*, 1986; L'Haridon *et al.*, 1993). A review of genotoxicity tests in amphibians is provided by Ferrier *et al.* (1998). DNA damage to invertebrates, especially marine invertebrates, can be assessed by a number of methods. In particular, tests of anaphase aberration have been done with polychaete worms and sea urchins, and reproductive success of invertebrates has been assessed in relation to genotoxicity (Anderson and Wild, 1994; Hose, 1994).

6.1 2 Teratogenicity

This second approach measures the incidence of deformities *(terata)* in newborn or newly hatched animals (reviewed in Hawkins *et al.*, 1995). The methods are similar to conventional single-species toxicity tests. For example, fertilized eggs of fathead minnows can be exposed until four days after hatching, with measurement of mortality and frequency of gross morphological deformities (lack of appendages, lack of normal fusiform body shape, and lack of mobility; USEPA, 1994b).

A similar test measures embryo and larval deformities in Japanese medaka. These are guppy-sized tropical fish that are quite sensitive to toxicants, and have large translucent eggs in which embryos can be observed easily. Eggs are exposed for up to two weeks, during development and hatching, and observed for lesions, deformation of the heart, failure of the swimbladder to inflate, and other developmental anomalies, as well as survival (Shi and Faustman, 1989; Marty *et al.*, 1991). Such a whole-organism test has obvious relevance to the environment, although the teratogenic effects might not be correlated with carcinogenesis. Adults of this species have been used in chronic laboratory tests.

6.1.3 Surveys of Tumours

A third approach surveys neoplastic lesions or developmental deformities among wild populations of animals. This is not a toxicity test but deserves brief mention as a tool for biological survey, possibly to check on predictions from laboratory tests. Liver neoplasms in bottom fish are the most commonly used, e.g., in English sole and related species at sites along the western U.S. coast, as an indication of sediment contaminated with PAH (Krahn *et al.*, 1984; 1986; Malins *et al.*, 1984; Overstreet, 1987; Stein *et al.*, 1992). The incidence of neoplasms tends to be much greater in heavily contaminated sites, but there are numerous exceptions, and correlations might not be seen with measured concentrations of PAHs. Neoplastic lesions on the skin and in the liver of brown bullhead, have been observed in a tributary of the Niagara River where sediments are contaminated with PAH (Black, 1983).

Neoplastic lesions and other histopathological changes are attractive biomarkers of genotoxic effects in the field. They are easily recognized and appear to be quite a sensitive reaction. However, a number of other factors can confound detection of neoplasms and interpretation of their incidence (tumours which do not represent genotoxicity, or simply age, diet, season, or reproductive cycle; Benson and Di Giullo, 1992). Even territorial fish tend to be mobile, weakening any associations of incidence and toxicants in the sediment (Krahn *et al.*, 1986). Reviews are provided by Mix (1986) and Harshbarger and Clark (1990).

6.2 Microscale Toxicity Tests

Key Guidance

- *Microscale tests use bacteria, unicellular algae, duckweed, rotifers, and other small plants and animals. Generally, culture is easy, and sensitivity of tests compares with that of larger organisms.*

- *A major advantage is the small size of samples. For testing effluents, cost and effort can be much reduced for collecting, shipping, and storing. Laboratory facilities can be much smaller. Microscale tests also reduce the need for testing larger, sentient animals. Their use widens the scope of test organisms, for better appreciation of effects in an ecosystem.*

- *Good microscale tests are fast, simple, reproducible, inexpensive, and require little space or time. Life cycles of the organisms are short, so inexpensive chronic and/or sublethal tests are feasible. They are useful for large-scale screening programmes.*

- *Some microscale tests have been condensed into kits that can be purchased off-the-shelf, need no culture of organisms, and are simple to run. Automation is a possibility for some tests. Procedures for quality control may be lacking.*

There are at least four good reasons for using very small organisms in toxicity testing. First, they reduce some major practical problems of handling large volumes of sample required by some conventional tests, thus reducing costs, sampling effort, shipping, storing and handling, and problems of laboratory space and tank size. All too often, investigators faced with these problems succumb to the temptation to run tests with less-than-optimal volumes, which lowers test quality and brings into question the accuracy of results.

A second clear benefit of moving towards very small organisms would be a reduction in testing on fish and other organisms which are complex and obviously sentient. This factor is becoming increasingly recognized.

The third advantage would be the encouragement of testing a wide range of types of organisms, not just fish, for example. All categories and types of organisms are important to the functioning of ecosystems, and for example, decomposing bacteria could well be key organisms for long-term maintenance of an ecological community. A bacterium might also be among the most sensitive organisms in a community. Small-scale tests with such organisms could be conveniently done with many replicates, improving precision of the results. A fourth benefit of using small organisms is their relatively rapid life cycle, allowing chronic tests to be completed during exposures of hours or a few days, i.e., a more direct estimate (and hence more accurate) of chronic sublethal effects.

Consequently, there is a strong trend in new tests toward "lower-order" organisms. Unicellular algae, bacteria, protozoans, small rotifers, roundworms, and similar creatures multiply rapidly and can be cultured in extremely small volumes of medium. Successful microscale tests tend to be reproducible and inexpensive, and encourage replication. They often allow efficient processing of many samples at once, and so they are useful for screening large numbers of samples (Blaise *et al.*, 1988; Dutka *et al.*, 1988). Some tests have been packaged into commercially available kits which bring additional advantages of portability, little or no culturing, and simplicity for use by non-specialists (Snell and Persoone *et al.*, 1989a,b). Microscale tests also have the benefit of lending themselves to automation, with potential savings in cost and in effort by investigators. For example, various automated adaptations of the Microtox test are commercially available (Qureshi *et al.*, 1998; Section 6.7).

A potential deficiency might be a lack of procedures for assurance and control of quality, since some of the methods are relatively new. Investigators should be aware of that possible weakness.

The most complete descriptions and assessments of microscale tests are provided in a book edited by Wells *et al.* (1998), and earlier comprehensive reviews are available (Bitton and Dutka, 1986; Dutka and Bitton, 1986). More specialized reviews are those of Blaise *et al.* (1988), discussing the utility of these tests in hazard assessment and monitoring by Environment Canada, and Dutka *et al.* (1988) who compared performance of a battery of microscale tests on water and sediments of the St. John River, New Brunswick.

As noted previously, there is an encouraging trend to test with various organisms, but this is almost always with a single species in each test. There is reasonably good indication that the results of such tests will predict the degree of effect in a real functioning community, but Cairns *et al.* (1998) point out that there is no absolute assurance of accurate prediction. They make a strong case for microscale community testing as a better predictor of effects on the interactions among species in real ecosystems. They point out that the tests with communities of microorganisms can now be done as economically as single-species tests. In validation experiments, such tests predicted overall ecological damage (or lack of damage) in 74% of the cases, a good record. Cairns *et al.* (1998) speculate that the failure of microscale community tests to achieve wide use, despite their ecological and technological validity, is because of a lack of charisma possessed by these humble organisms compared to such popular symbols as "bald eagles or sea otters". This point seems valid in view of the continuing emphasis on tests with rainbow trout for legal purposes, rather than other taxonomic components of ecosystems.

6.2.1 Examples of Tests Available

This section mentions a few of the tests which are commonly used in Canada or appear to be coming into routine use. There is a numerous and varied choice of other microscale tests available, as may be ascertained by consulting the large book on such tests (Wells *et al.*, 1998). Some of the other techniques covered in that book, which could be of interest to investigators, include tests at the biochemical level, including immunoassays (Dankwardt *et al.*, 1998), submitochondrial particles (Read *et al.*, 1998), enzymes such as acetylcholinesterase (Obst *et al.*, 1998), and tests with cultures of tissue from various organisms (Müller and Müller, 1998; Denizeau, 1998; Durand *et al.*, 1998). The book also has large sections with chapters on various kinds of organisms: bacteria, algae, protozoa, invertebrates, and vertebrates.

Two microscale tests that are routinely used are the **bacterial luminescence test** (EC, 1992c), and the microplate version of the growth inhibition test with a **freshwater alga** *(Selenastrum capricornutum;* EC, 1992d; Blaise *et al.*, 1998). The luminescence test uses the bacterium *Vibrio fischeri* (formerly *Photobacterium phosphoreum)*, commonly by using the Microtox test, and its development and adaptations are described in Qureshi *et al.* (1998) and Johnson (1998). Variations of the algal test continue to develop, and one that is available commercially, provides immobilized alga so that it is not necessary to maintain a culture in the laboratory (Persoone, 1998a). Common microscale genotoxicity tests are described in Section 6.1.

A spectrum of other microscale tests are available or under development, many having the advantage of sublethal endpoints. The **ATP-TOX system** (Xu and Dutka, 1987; Xu, 1998) and the **algal ATP** toxicant screening test (Blaise *et al.*, 1984) are similar tests based on inhibition of cell growth and enzyme activity. They measure production of the enzyme ATP in cultures of a bacterium or green alga. ATP is measured by light production in the luciferin-luciferase system (the enzyme system that produces light in fireflies) as a percent of controls. Thus, it measures numbers of organisms and hence the rate of cell multiplication from the starting point.

The **Toxi-Chromotest** is a rapid method for water, similar to the SOS-Chromotest (Section 6.1.1) in measuring production of the inducible enzyme ß-galactosidase by the bacterium *Escherischia coli*. Stressed bacterial cultures are exposed to a known inducer of ß-galactosidase and a substrate for the enzyme, then exposed to test solutions in microplate wells. Toxic materials interfere with recovery of the cells from stress and thus reduce ß-galactosidase synthesis, detected by a lack of colour change from a chromogenic compound linked to the substrate. As with the SOS-Chromotest, use of a colorimetric plate reader is recommended to improve accuracy.

The **Thamnotoxkit F** and the **IQ test** with ***Daphnia magna*** are two commercially available microscale tests which proved to have the best sensitivity to various kinds of mining wastes in an assessment under the AETE program of the Canadian government and mining industry (Pollutech, 1996). The Thamnotoxkit F is a 24-h lethal test performed in a multi-well test plate using an early larval stage of a fairy shrimp *(Thamnocephalus platyurus)*. Conveniently, there is no need to maintain a stock culture in the laboratory since the organisms are held as dormant eggs *(cysts)* and reared for 24 h to obtain larvae for the test. The results of this test showed food correlation coefficients of 0.884 to 0.92 with the standard lethal test using *Daphnia magna*, and a coefficient of variation of 13.5% was obtained in a series of tests on a reference toxicant, conducted by the same person (Persoone, 1998b). The background and rationale for this test and similar ones is also provided by Persoone *(loc. cit.)*. The IQ test uses young *D. magna* which have been starved. They are apportioned to the test concentrations and provided with fixed amounts of the substrate methylumbelliferyl-ß-D-galactoside *(MUF-galactoside)*. Viable *Daphnia* ingest the substrate and hydrolyze it enzymatically, producing the substance 4-methylumbelliferone which is fluorescent (the process is described in Lee and Tay, 1998). Failure to carry out the hydrolysis is considered a toxic response. Individual daphnids are categorized as fluorescent (i.e., viable) or not, and the EC50 is calculated. The test is not well documented in the scientific literature at the time of writing, but its performance with mining waste is evaluated by Pollutech (1996), and it is commercially available (Aqua Survey Inc., Flemington, N.J., USA). The **Ceriofast™** test measures suppression of feeding by *Ceriodaphnia dubia* in a 1-h exposure, and results were well correlated with the standard 48-h test with *C. dubia* (Jung and Bitton, 1997).

A promising **multi-species algal test** simultaneously exposes five algae in 96-well microplates for two or three days. Growth and reproduction is estimated as the number of cells by almost instantaneous measurements of fluorescence using a special device. Use of the five species gives excellent sensitivity to diverse categories of substances (SRC, 1995a).

A simple test measuring the 24-h LC50 for **freshwater or marine rotifers** has been extensively tested in Europe, showing sensitivity similar to daphnids (Persoone *et al.*, 1989; Snell and Persoone, 1989a,b). Animals can be stored for a year as dormant cysts, eliminating cultures. They are hatched by a temperature change, and the active hatchlings are tested. **A large aquatic bacterium** (*Spirillum volutans*) is used in another test, to assess loss of motility after a 2-h exposure. The motility is normally accomplished by a rotating fascicle of flagella at each pole of the bacterium, which can be observed.

6.3 Tests with Microcosms and Mesocosms

Key Guidance

- *A true microcosm not only contains a variety of species, but must also reflect the behaviour of an ecosystem. It can be held in a flask, box, or aquarium in the laboratory, where confounding environmental variables can be controlled.*
- *A toxicity test with microcosms can be revealing because it tests several species, their interactions, allows movements and transformations of test chemicals, and toxicity of metabolites.*
- *Microcosm tests are sometimes more sensitive than single-species ones, but are usually comparable. Population dynamics and species interactions can dampen sensitivity of the tests.*
- *Microcosms tend to be variable, making replication difficult. Standard methods have not yet been developed for routine testing. Microcosm tests are more demanding of space, time, funds, and expertise than single-species tests.*
- *Mesocosms are a useful research tool, but tests with these large enclosures are too expensive and time-consuming for routine use in testing or monitoring.*
- *Aquatic mesocosms could be recirculating artificial streams in a building, or natural stream channels some hundreds of metres long, with continual input of new water and test chemical. Those mesocosms are relatively stable, successful in replication, and have been useful in validating results of single-species tests.*
- *Mesocosms could also be sections of land, ponds, bag-type enclosures in the ocean, or even small lakes. The static aquatic mesocosms can show large variation in replicates, and meaningful analysis of their (usually) declining test concentration remains a question.*

Toxicity tests with multispecies assemblages are intended to address interaction of species and the consequent secondary effects. They should provide more valid predictions from controlled laboratory tests to the effects of a pollutant in nature. Unfortunately, they have not reached that combination of reliability, ease, and economy that would make them suitable for routine use.

The term "microcosm" has the connotation of a miniature world or ecosystem, in fact a simplified community maintained in a laboratory or semi-natural situation. *Microcosms* are relatively small systems. An aquatic microcosm might have a maximum volume of 3 or 4 litres, and contain a dozen or fewer species of microorganisms including small crustaceans. (Some definitions have outdoor aquatic microcosms as large as 15 m^3.) Usually the whole assemblage is contained in a flask or other container in the laboratory. *Mesocosms* might range from the size of an aquarium upwards to a few tens of square metres of land, sections of ponds, or ponds themselves.

6.3.1 Tests with Microcosms

Tests using controlled mini-communities would seem ideal for meaningful appraisals of toxicity, but at present, they are mainly used at the research level. Microcosms are a long way from gaining a role in routine testing because of the degree of effort and cost, and variability in performance of the small communities (see the following text).

Information on toxicity tests with microcosms is given by Cairns (1985), Buikema and Voshell (1993), and Graney *et al.* (1995), and on multi-species tests by Kimball and Levin (1985). Criteria for applicability of results to nature are discussed by Ausmus and O'Neill (1978, dealing specifically with solid systems) and in various papers edited by Levin and Kimball (1984). Guidance for creating soil-core microcosms is available (Van Voris *et al.*, 1985; ASTM, 1997a).

Description. These test systems are variously called multi-species tests, micro-ecosystems, or microcosms. Technically, any system in which several species are exposed to the toxicant simultaneously is a multi-species test (Cairns, 1985). A true microcosm is a simplified analogue of a natural ecosystem, designed to reflect the ecosytem's trophic levels, structure, and behaviour, but housed in a laboratory in order to gain some control of confounding environmental variables (Kersting, 1984; Taylor and Parkinson, 1988). The microcosm must incorporate key ecological processes and allow for population fluctuations and community interactions (Van Voris *et al.*, 1985). In practice, those requirements often go unrecognized, so that most multi-species systems are called microcosms.

A wide variety of microcosms have been developed for testing toxicity in aquatic and terrestrial systems. Aquatic systems include static, recirculating, and sometimes flow-through systems patterned after lakes, wetlands, or rivers. Marine microcosms are less common. Terrestrial microcosms are usually soil subsystems. Most microcosms consist of a container (flask, aquarium, box) with water and sediment, or soil that is usually taken from the environment. Organisms are added either as natural assemblages from the field, or by introducing selected species. Microcosms based on soil cores already contain a full complement of organisms (ASTM, 1997a). Rock-filled trays have been set in a stream, allowed to colonize with invertebrates for 30 d, then transferred to recirculating streams in the laboratory (Clements *et al.*, 1988; Pontasch *et al.*, 1989). A standard aquatic microcosm for testing pesticides (Taub, 1985) consists of specific sediments, water with added nutrients, 10 species of algae, a rotifer, a protozoan, and three small crustaceans (*Daphnia*, *Hyalella azteca* and a seed shrimp, *Cypridopsis*).

There is a wide latitude in manner of introducing the toxicant. There can be a low-level chronic concentration, a single pulse at high concentration, or several pulses. Endpoints include changes in density or biomass of species or groups of species (e.g., soil invertebrates) and changes in community structure (species richness, similarity indices, predator-prey interactions, and proportions of different groups or other measurements of dominance). In a complex microcosm, those variables could be measured for each trophic level, taxonomic group or habitat (e.g., water and sediments). Community-level processes are sometimes measured as well, particularly primary and secondary production, decomposition, and community metabolism (Buikema and Voshell, 1993). Indigenous microbial communities in samples of surface water were assessed for their ability to oxidize 95 different sources of carbon, in a laboratory test that is claimed to be fast and economical (Lehman *et al.*, 1997).

Advantages and disadvantages. Microcosm tests have many potential advantages over single-species toxicity tests. In addition to direct toxicity to a number of species, microcosms can assess some of the movements, transformations, and fates of toxic chemicals, and also any toxicity arising from metabolites. They can be controlled, manipulated, and replicated. Reasonably realistic microcosms can simulate responses of whole natural ecosystems (Buikema and Voshell, 1993), and applicability of results to nature depends on that realistic imitation, particularly the interaction of species.

Sensitivity to toxicants does not seem to be particularly high for microcosm tests, and that removes some of the impetus for routine use. Effects on community structure have sometimes been observed at levels well below the most sensitive result for single-species tests (Dewey, 1986; Pontasch *et al.*, 1989). However, greater sensitivity of microcosms is observed less often than might be expected, given the greater likelihood of one or more sensitive species (Kooijman, 1985). A few comparisons between the thresholds of effect predicted from single-species tests, and the observed thresholds from microcosm tests, have generally confirmed the applicability of the single-species approach (Pontasch *et al.*, 1989; see Section 4.1.2). The most important factor contributing to the lack of enhanced sensitivity seems to be that toxicity to individuals is not always reflected in toxicity to populations, and population interactions tend to further dampen responses at the community level. Some microcosm studies measure only lumped variables like biomass of all algae, and thereby lose sensitivity by missing effects on species (Kooijman, 1985).

On the negative side, microcosms are far more demanding of space, time, funds, and expertise than single-species tests (Mothes-Wagner *et al.*, 1992). One of the greatest problems in microcosm tests is the difficulty in reproducing results because the complex micro-communities vary from one microcosm to the next, even for replicates. It is often impossible to prevent extinctions and changes in community structure, irrespective of the toxicant, especially in longer tests (Buikema and Voshell, 1993). There are also problems of scale, often created by the lack of habitat variation, high ratio of surface area to volume ("edge effects") and the container ("wall effects") (Giesy and Allred, 1985). For example, a single predator per container might represent a very high density compared to nature, and it might eradicate a prey population if there were no refuges. Toxicant-related differences tend to be obscured because of the difficulty in reproducing results (Kersting, 1984). Standardized artificial communities tend to be less variable (Giesy and Allred, 1985), but standardization limits the realism for a given local problem (Pontasch *et al.*, 1989).

Every microcosm or multi-species test represents a compromise between precision (reproducibility of data), and accuracy (agreement with the real world). Microcosms are models of natural ecosystems and involve much simplification (Giesy and Allred, 1985). The need to verify microcosm results in natural ecosystems has been stressed (Cairns, 1986; 1988), but that is difficult and the criteria for verification remain uncertain (Buikema and Voshell, 1993).

6.3.2 Tests with Mesocosms

Mesocosm testing will undoubtedly continue its role as a research tool, not a routine testing method, because of the size of these operations and their duration of weeks, months, or occasionally years. Still, some review is warranted here, partly because of the excellent performance of mesocosms in validation of single-species tests. Mesocosms are fairly widely used in aquatic studies, and to a lesser extent in soil research. Detailed review of aquatic mesocosms is provided by Graney *et al.* (1994; 1995), and examples of their use are given in Section 4.1.2.

Description. An example of a small aquatic mesocosm would be recirculating artificial streams of the order of 6 m length, operating in a laboratory. They can test small but complete and realistic communities containing micro-organisms, algae, invertebrates and a few fish (Warren and Davis, 1971).

Most aquatic mesocosms are outdoors and some colonization might be allowed from surrounding natural areas. A very successful design for toxicity testing in fresh water uses a series of parallel, natural-substrate artificial streams constructed outdoors. In such a facility at Monticello, Minnesota, stream channels are of realistic size (1.4 m wide in riffle areas and widening in pools), replicates can be used, and steady flows of new water can be dosed to constant concentrations (Zischke *et al.*, 1983; Cairns, 1985). Small populations of fish find a suitable niche, and their survival, growth, and reproductive success under the semi-natural conditions can be measured during an experiment of a few months. Other components of the community can be evaluated by standard survey techniques or by removing blocks of artificial substrate.

Several groups of aquatic investigators have studied large volumes of lake or seawater enclosed in plastic bags or in "limnocorrals" running to the bottom, with toxicant added to some of the enclosures. Fish are usually excluded, and populations of other organisms are assessed by conventional sampling techniques. Enclosures are reasonable research tools but share certain difficulties with other types of closed microcosms and mesocosms. At the upper extreme of manipulated communities that might be called mesocosms, small lakes have been dosed and used as experimental systems, as in the Experimental Lakes Area of northwestern Ontario, where some environmental toxicology has been done by the Canadian Department of Fisheries (Schindler, 1987).

Advantages and disadvantages. Among aquatic mesocosms, the outdoor artificial streams mentioned above have been relatively effective. Their near-natural communities tend to be stable, and replication has been relatively successful. One of the biggest toxicological contributions of aquatic mesocosms has been their use in testing the validity of single-species tests, as a basis for establishing water quality guidelines. Again, outdoor artificial streams have been productive in this regard, and have given reasonable validation of single-species tests (see Section 4.1.2).

One of the biggest disadvantages of mesocosms is the cost. An experiment involving half-a-dozen people for a summer is relatively expensive.

Replication is not usually a severe problem in laboratory streams or outdoor stream channels, but it is in many static systems. As in microcosms, segregated pond communities tend to be unstable, and replicates might diverge in characteristics for reasons that are not obvious. An example is seen in twelve ponds of 20 m on a side, which developed somewhat similar patterns of communities during a year, but had different community structures that could not be used as replicated test systems (Rosenzweig and Buikema, 1994). Closing off a section of a natural or artificial pond can allow better control and replication.

Another common difficulty in studies with static aquatic mesocosms is that constant concentrations are not maintained because toxicant is usually added only once. The declining concentration of toxicant is not comparable with the usual laboratory tests and is difficult to interpret in a meaningful way because it is confounded with duration of exposure. It is difficult but not impossible to maintain a constant concentration in enclosure-type systems (Kuiper, 1981).

6.4 Body Burdens

Key Guidance

- *Lethal effects of neutral narcotic chemicals for fish occur when the total body burden of all such organic chemicals reaches 2–6 mmol/kg, regardless of the chemical concentrations in the water.*
- *Comprehensive documentation has not yet been obtained for critical body burdens of other groups of chemicals, and models for groups of organisms other than fish is also incomplete at present.*

Endpoints from aquatic toxicity tests are normally expressed in terms of the concentration in the external medium. This makes sense from a regulatory point of view because concentrations in the water are measured in routine monitoring, and it is the concentration in effluents and other wastes that can be controlled by treatment. However, it is the concentration or amount of toxicant at the site of toxic action within the organism (the true *dose*) that determines whether the response is observed, regardless of the external concentration (McCarty, 1991). Since the amount of toxicant at the site of action is difficult to measure, and even the site might not be known, then a reasonable substitute is measurement of whole-body residues of the toxicant. Internal dose-response relationships could describe the dynamics of toxic action more accurately, allowing better scientific prediction, particularly for such things as toxicity of mixtures of substances (McCarty, 1991).

It is clear that toxic effects of neutral narcotic[56] chemicals on fish occur when a critical concentration is reached within the organism (McCarty, 1991; McCarty *et al.*,

---

[56] This does not refer to illegal drugs used by humans, but to organic compounds with a non-specific action, which hamper and destroy the general functions of cells.

1992; McCarty and Mackay, 1993). This has been tested many times, and toxicity can be predicted readily from regression equations (QSARs) based on the octanol–water partitioning coefficient (log K_{ow}) of the compound, which in turn estimates how readily the substance enters the body (see Section 4.6). A critical body burden in the range of 2 to 6 mmol/kg is predicted for moribund fish in lethal tests (McCarty, 1991), and published measurements support the predictions. Estimated lethal body burdens vary by a factor of only three to four, for substances which have LC50s ranging over five orders of magnitude (McCarty, 1991; McCarty *et al.*, 1992).

For sublethal toxicity, QSAR suggests that critical body burdens should be roughly 1/10 those for acute lethality, but so far there are too few data to test this prediction (McCarty and Mackay, 1993). In theory, the same principle should apply to other modes of toxic action. Critical body burdens might be very different for chemicals acting by different mechanisms, but should be similar within a group. McCarty and Mackay (1993) have estimated critical body burdens in fish for each of the eight known classes of toxic action for organic compounds, based on preliminary data. Their analysis indicates distinct body burdens for different classes of toxic action, but data within some classes span two or three orders of magnitude.

Outside the group of narcotic organic chemicals, more research is required before the body burden approach can be applied to general categories of substances. For example, the metals tend to have specific modes of action, so they might have various critical body burdens, probably very different for micronutrients (e.g., copper) and metals that are not micronutrients (e.g., lead) (McCarty and Mackay, 1993). The concept of a critical body burden does, however, fit the observations for individual metals. For example, it has long been thought that the action of hard water in lessening toxicity of metals to aquatic organisms occurred because the hardness cations slowed passage of metals through membranes into the organism, thus delaying and reducing accumulation in the tissues. Experiments by Borgmann *et al.* (1991) support that view. The sublethal critical body burden of cadmium for a freshwater crustacean *(Hyalella azteca)* changed by a factor of only 2, while the concentration in the water that was toxic changed 10-fold with water hardness, and a 1000-fold with addition of sorbing sediments. Critical body burdens changed only 16% when chelating compounds caused a 35-fold increase in the waterborne cadmium necessary for toxicity. The picture might not be simple, however, since critical body burden of arsenic for trout is temperature-dependent, possibly because the dominant form of the metal changes with temperature (McGeachy and Dixon, 1990).

Associated avenues of investigation are being explored, such as dietary accumulation of toxicants as a route for attaining critical body burdens. Some quantifications have been done for fish, but the overall relationships are not yet clear (Fisk *et al.*, 1998). Some success has been achieved in *predicting* body burdens of chemicals in a complex mixture, rather than attempting to measure them in the organism by chemical analysis. An advantage is that the predictive method can include substances not likely to be detected in direct chemical measurements of tissues, because of their low concentrations. The procedure subjects the complex mixture to bioconcentration into a hydrophobic phase, then concentrations in that phase are measured by advanced chemical methods. The authors conclude that "using these estimated total body burdens, baseline toxicity effects can be predicted, including the contributions of chemicals with specific modes of action to the overall baseline toxicity" (Van Loon *et al.*, 1997).

Work with eggs and larvae of fish indicates somewhat greater uptake of PAHs and PCBs than predicted by K_{ow}, suggesting that octanol is good, but not ideal, for representing fish lipids (Petersen and Kristennsen, 1998). Continuing research shows that the relationships can be improved. There have been differences observed between lethal concentrations and lethal body burdens, for chemicals of the categories *Narcosis I* (inert narcotics such as butanol and xylene) and *Narcosis II* (polar narcotics such aliphatic and aromatic amines and phenols). These can be resolved by basing the QSAR on the partition coefficient between water and, instead of octanol, the membrane surrogate L-α-dimyristoyl phosphatidylcholine (Vaes *et al.*, 1998). Modes of action have become a focus of efforts. Russom *et al.* (1997) combined a variety of information from toxicological fields to develop procedures for relating modes of lethal action in fathead minnows to chemical structures and properties. They classified approximately 600 substances into eight categories of toxic action: three groups of narcotics; oxidative phosphorylation uncouplers; respiratory inhibitors; electrophiles/proelectrophiles;

acetylcholinesterase inhibitors; and seizure agents for the central nervous system. A computer-based system was established, to associate chemical structure with likely mode of toxic action, and corresponding QSAR if available.

The literature on other organisms cannot be completely reviewed here, but a few examples illustrate the various themes. Often the focus is different from the critical body burden concepts for fish, but sometimes there are parallels. Data on critical body burdens of earthworms exposed to two chlorobenzenes agreed with measured data for fish (Belfroid *et al.*, 1993). Body residues of pentachlorophenol in earthworms have been applied to assess the effects of modifying factors on toxicity (Fitzgerald *et al.*, 1996). Encouraging results were obtained for critical body burden of fluoranthene (a polycyclic aromatic hydrocarbon) in amphipods living in the sediment. The critical burdens appeared to be generally similar to the range for fish, at about one to somewhat more than 6 μmol/g of tissue.

For green plants, uptake from the soil was well described by a QSAR called the molecular connectivity index, which is related to size of molecules, their branching, molar volume, and surface area. Predictions were somewhat superior to those based on octanol–water or octanol–air coefficients (Dowdy and McKone, 1997). Often the emphasis is on residues *on* plants, rather than residues solely within the tissues, e.g., for biomonitoring sentinel plants to assess atmospheric deposition of herbicide residues (Felsot *et al.*, 1996). For pesticides a particular food-chain model (the Kenaga nomogram) has been developed by the USEPA, as an instrument for estimating pesticide residues on plants (Fletcher *et al.*, 1994); the model also has a function in estimating intake of pesticides by animals which eat the plants (Pfleeger *et al.*, 1996).

Among vertebrates, tissue concentrations of toxicants have received a great deal of study over many decades, but usually on a toxicant-by-toxicant basis rather than on any predictive models for classes of compounds. Frogs in Ontario were studied for accumulation of organochlorine residues (Russell *et al.*, 1995). Tissue residues in freshwater turtles have been assessed as a means of monitoring chemical and radioactive contamination (Meyers-Schöne *et al.*, 1993).

Contaminants in the tissues of birds, and more particularly in their eggs, have been correlated with apparent toxic effects in many studies, for example egg mortality among double-crested cormorants of the Great Lakes, associated with residues of polychlorinated biphenyls. Food-chain uptake is also studied in birds (Forsyth *et al.*, 1994).

6.5 Mixed-function Oxidases

Key Guidance

- *Mixed-function oxidases (MFOs) are enzymes in animals that detoxify a number of natural and anthropogenic substances.*
- *Elevated levels of MFO activity have been demonstrated in fish, for a wide variety of sites contaminated by PAH, PCB, or oil products.*
- *The strength of MFO induction in fish is affected by species, season, sex, age, temperature, feeding, and especially reproductive status, sometimes strongly enough to obscure effects of pollution.*
- *Induction of MFO enzymes indicates exposure of fish to inducers; the significance to the health of the fish is not known.*

Many organisms detoxify lipophilic substances by means of membrane-bound enzymes known as mixed-function oxidases (MFOs). They catalyze the conversion of toxicants into water-soluble forms that can be excreted (Addison, 1984)[57]. In nature, MFOs function in the metabolism of steroids, vitamins and some bile acids. They are useful as bioindicators because they are inducible—their activity increases in the presence of contaminants[58]. MFOs are found in many different organs but in vertebrates, they are concentrated in the liver. Hodson *et al.* (1998) have provided an authoritative and complete review of MFO use in assessing effluents.

---

[57] MFOs are associated with iron-containing hemoproteins, cytochromes P-450 (so-called for the wavelength of their absorption peaks) as terminal oxidases (Payne *et al.*, 1987).

[58] The enzymatic action of MFOs can be measured by the accumulation of metabolites. Activity of liver mono-oxidases is usually measured by extracting the enzymes and providing a supply of a substrate known to induce their action, such as benzo(*a*)pyrene. The rate of accumulation of metabolized product is then measured. Three functionally defined enzymes are routinely assayed, according to the substrate used in the assay (Stegeman *et al.*, 1988): 1. aryl hydrocarbon hydroxylase (AHH), sometimes called BaP-ase; 2. ethoxyresorufin *o*-de-

In fish, the MFO system is induced by a relatively narrow range of compounds. Elevated MFOs have been demonstrated at various sites contaminated by PAHs, PCBs, or oil products (Addison, 1984; Payne *et al.*, 1987) or in laboratory aquaria dosed with oil (e.g., Payne, 1978; Payne and Fancey, 1982). Effluents from pulp mills, municipal treatment plants, and storm-sewers can be weak inducers (Payne *et al.*, 1987; Andersson *et al.*, 1988; Hodson *et al.*, 1991). In particular, induction has been caused in the past by dioxins and other components of effluents from pulp mills which bleached strongly with chlorine (Ahokas *et al.*, 1994). Sources in the mill processes include cooking liquor, black liquor, and effluents from the first stage of bleaching (Hodson *et al.*, 1997), and secondary treatment of effluent can reduce or eliminate MFO induction (Martel and Kovacs, 1997). At one mill, the causes of MFO induction were found to be two juvabiones, natural components of balsam fir wood (Martel *et al.*, 1997). The reaction has been used to evaluate the apparent potency of polychlorinated dibenzo-*p*-dioxins and dibenzofurans (Parrott and Hodson, 1995). Theoretically, MFO induction is proportional to the level of environmental contamination, taking into account the status of the fish and the bioavailability of the contaminants.

The significance of MFO induction to the health of the fish is not known (Payne *et al.*, 1987; Kleinow *et al.*, 1987; Hodson *et al.*, 1991). On the one hand, a high rate of MFO activity could lead to increased tolerance of pollution because of rapid elimination of toxic compounds (Stegeman and Kloepper-Sams, 1987). On the other hand, detoxification of PAH can produce metabolites that are known carcinogens. The few studies on relationships between MFO induction and fish health have demonstrated correlations in some cases (Spies and Rice, 1988) but not in others (Stegeman *et al.*, 1988). A question asked about development of MFOs by birds is whether this should be considered biochemical acclimatization or adaptation (Fossi *et al.*, 1991).

At present, MFO induction can serve as an early warning of exposure to PAHs, PCBs, and some effluents, and has been used as an exploratory tool for detecting problems, e.g., by testing sturgeon in the St. Lawrence river (Rousseaux *et al.*, 1995). For early warning, measurement of MFOs is most effective when combined with other biomarkers, to overcome variability (Payne *et al.*, 1987; Jiminez *et al.*, 1990; Stein *et al.*, 1992). A non-destructive technique that might prove useful in laboratory testing or surveys measures fluorescence in the bladder of embryonic fish. The EROD substrate ethoxyresorufin, is added to test water and in the presence of appropriate toxicants, fluorescence occurs and can be measured with suitable precision (Nacci *et al.*, 1998).

There is a base variation of MFO activity in fish, resulting from various physiological and environmental factors. Species, season, sex, age, temperature, and feeding status can all affect MFO induction, sometimes strongly enough to obscure site differences (Hodson *et al.*, 1991; Kleinow *et al.*, 1987). Reproductive status is the most important biological factor influencing MFO induction (Hodson *et al.*, 1991). There are also marked seasonal patterns in MFO at clean sites. Low temperatures tend to delay and suppress MFO induction. Feeding status can also moderate MFO response, although this in turn is often related to the reproductive cycle (Jiminez and Burtis, 1988; Ankley and Reinert, 1988).

Because of this variation in the base level of MFO activity, MFO induction at a polluted site also tends to be variable, and gradients of effect can be difficult to distinguish. Despite the abundant literature, few experiments relate MFO activity to the range of contamination in sediments (examples in Payne *et al.*, 1987; Stegeman *et al.*, 1988; Addison and Edwards, 1988). There appear to be relationships between environmental levels of inducers and MFO activity, although they are frequently obscured by variability and non-linearity of response.

Equally important as the variation, for obtaining good results, is the choice of a reference population of organisms, since MFO activity is evaluated by the difference between and exposed and reference group of organisms. Considerable care must be taken to choose a comparable reference group, and sometimes that is not

---

ethylase (EROD); and 3. 7-ethoxycoumarin *o*-de-ethylase (ECOD). Sometimes the amount of cytochrome P-450 is used as an MFO assay, but it is less sensitive than enzyme activity (Payne *et al.*, 1987).

entirely feasible. Because of environmental influences on MFO activity at the same site (previous paragraph) caution must prevail in making historical comparisons.

Studies on fish have been outlined in this section, but other organisms have similar responses involving MFO. Mouse cells cultured in the laboratory have given a sensitive response of an MFO (EROD, ethoxyresorufin *o*-de-ethylase), providing a sensitive bioassay for various elements of pulp mill effluents (Koistinen *et al.*, 1998). Measurements of MFOs have been made in many other organisms, from mussels (Solé *et al.*, 1995) to birds such as starlings (Trust *et al.*, 1994), and gulls living in environments affected by humans (Fossi *et al.*, 1991).

6.6 Endocrine Disrupters

Key Guidance

- *Interference with the reproductive physiology of fishes was apparent in Canadian studies near bleached kraft pulp mills in the late 1980s and early 1990s. More recent work at various locations indicates complex situations rather than a simple cause-and-effect relationship.*
- *Sexual abnormalities have also been documented in other kinds of wildlife, in some cases clearly or apparently related to pollution by organochlorine substances.*
- *Other examples of endocrine disruption involve organotin and snails, and cyanide and fish.*
- *The topic has received wide public and scientific attention, and reviews and books are available.*

Effect on reproductive biochemistry of wildlife has become a major focus of ecotoxicology in the 1990s. This kind of effect from the action of certain toxicants is of particular interest in Canada because components of pulp mill effluent apparently cause the effect in aquatic organisms. Swedish research in the Baltic during the 1980s suggested reproductive effects on fish in locations near bleached kraft mills (Södergren, 1987; Södergren *et al.*, 1988). Studies commencing in 1988 at mills in northern Ontario indicated that exposed fish had decreased levels of serum steroids including testosterone and estradiol, with consequent abnormalities in reproductive behaviour and carbohydrate metabolism (McMaster *et al.*, 1991; Munkittrick *et al.*, 1991). The fish grew more slowly than fish from reference sites, matured more slowly, had smaller gonads, reduction in secondary sex characteristics in males, failure of females to develop eggs of increased size, and other apparent effects. Since then, similar effects have been documented elsewhere in Canada (reviewed in Kovacs *et al.*, 1996).

There were some unexpected findings, such as an apparent lack of effect on fertilization potential of fish exposed to the mill effluent, and no effect on hatchability of eggs or size of the hatched larvae (McMaster *et al.*, 1992). Furthermore, installation of secondary treatment at a bleached kraft mill did not improve the levels of reproductive steroids, nor did two weeks of shutdown of mill operations (Munkittrick *et al.*, 1992). Studies at another mill with secondary treatment did not show any conclusive reduction of steroid levels (Servos *et al.*, 1992). Studies in a Quebec river led the investigators to conclude that effects on fish gonads apparently resulted from natural causes, rather than from exposure to bleached kraft mill effluent (Gagnon *et al.*, 1994).

Clearly the toxicological picture was complex. A large amount of research was generated, particularly in Canada, and a thoughtful review of findings to the mid-1990s is provided by Kovacs *et al.* (1996). Research has continued, with documentation that mill effluent reduces the capacity of fish to synthesize reproductive steroids (McMaster *et al.*, 1996). Attempts have been made to identify the substances that cause the effects, with some indication of activity by natural components of wood used in pulping. A review of the relevant endocrine physiology of fish has been provided by Arcand-Hoy and Benson (1998).

Some parallel studies have been done on other kinds of organisms. In Canada, poor reproductive performance of great blue herons and eagles might have been related to polychlorinated dibenzodioxins and/or other components of bleached kraft effluents, although there could have been various other causes (Elliott *et al.*,1989; Elliott and Norstrom, 1998). Perhaps the most striking example has been gross effects on the reproductive development of alligators in Lake Apopka, Florida, apparently the result of heavy pollution by organochlorine compounds (Guillette *et al.*, 1996). Reviews are available for the mechanisms and effects of endocrine disruption in mammals and other wildlife (Gray *et al.*, 1998; Campbell and Hutchinson, 1998).

Since about 1994, the subject of endocrine disruption has gained wide public and scientific attention. Substances other than organochlorines have been associated with this phenomenon, and one of the most widely known cases has been masculinization of sex organs in marine snails, associated with the use of tributyltin in antifouling paints on boats (reviewed by Matthiessen and Gibbs, 1998). Effects in fish have been noted following exposure to cyanide (Ruby *et al.*, 1993). Many publications have ensued. Reviews are available on mechanisms of steroid-like actions of contaminants (Gillesby and Zacharewski, 1998) and screening methods for detection of such active chemicals (Ankley **et al.**, 1998). Several books are also available. Keith (1997) describes the properties of 67 known or suspected endocrine disrupters. Mechanisms in humans (Naz, 1998) might have relevance to environmental toxicology for mammalian wildlife. Kendall *et al.* (1998) focus on wildlife including mammals, birds, and fish, and Rolland *et al.* (1997) deal specifically with fishes.

Knowledge concerning this spectrum of effects is far from complete, but already the findings have served as a potent force for reducing input of waste chemicals to the environment.

6.7 Continuous On-line Biomonitoring

Key Guidance

- *Continuous physicochemical or biomonitoring of an effluent can alert operators to an unexpected surge of waste, with no time lag and the possibility of avoiding damage in receiving waters.*
- *Most on-line biomonitoring uses respiratory movements of fish, but various other endpoints have been used or attempted.*
- *Biomonitoring systems require sophisticated electronic equipment, and statistical analysis of the time-series data is unstandardized and difficult. Responses of organisms can vary with time of day, season, or physiological status.*
- *On-line biomonitoring has seen some use in industry.*

Effluents are routinely monitored by sampling for chemical analyses and toxicity testing. The results might not be available for several days, however, by which time it is too late to stop any unexpected peak in discharge. A pulse of toxicity lasting a few hours, or even days, might be missed by normal sampling schedules. Continuous monitoring can overcome some of the limitations of periodic sampling. On-line physicochemical measurements (e.g., pH, conductivity, certain chemicals) can sample nearly continuously and sound alarms when there are excursions in effluent characteristics. Corrective actions might be possible before there was significant escape of effluent to the receiving system. Continuous biomonitoring systems could function similarly, with the usual complementary advantages of biological endpoints.

In a continuous biological monitoring system, a portion of the effluent stream is diverted through holding tanks in which test organisms are maintained. The well-being of the organisms is judged by some aspect of their behaviour or physiological performance, and is continuously monitored electronically. Deterioration of water quality is signalled by change in behaviour or performance, which can be detected very quickly with computer-assisted analysis, and an alarm can be triggered. That is the optimistic scenario, and a number of biomonitoring systems have been designed with varying degrees of success.

To date, most on-line biomonitoring has relied on fish. The most frequent endpoints are body movements associated with respiration, especially rates of gill irrigation and "coughing" (Carlson and Drummond, 1978; Thompson *et al.*, 1982; Fisher *et al.*, 1983; Westlake *et al.*, 1983; van der Schalie, 1986). Increases in irrigation and coughing rates have long been viewed as an indicator of stress (see citations in Thompson *et al.*, 1982).

Many other organisms have been used, including algae (van Hoof *et al.*, 1992; Pandard *et al.*, 1993) and benthic invertebrates (Coler *et al.*, 1988) but for the most part these systems are still experimental. On-line systems with bacteria use an automated version of the luminescence assay (Microbics, 1992; Qureshi *et al.*, 1998). In one version, a computer-controlled manipulator reconstitutes the freeze-dried cultures, extracts and mixes the test samples, and transfers cuvettes to a photometer for measurement of light emission. The computer analyzes the data and can be programmed to sound an alarm when light production declines. The system is not strictly continuous, but does hourly tests. Some other responses and organisms that have been used are mentioned in the following discussion.

The invention of sensitive electronic sensors has revolutionized biomonitoring systems (Pandard *et al.*, 1993). Gill irrigation can be measured by electrodes mounted on the walls of the tank to detect the minute bioelectrical signal generated by contraction of muscles in the pharyngeal region of a fish. The signals are amplified, accumulated, and stored on computer for analysis. Cairns *et al.* (1980) describe such a system in detail, and much more sophisticated sensors and software are available today (see references in Isom, 1991). Biosensors can measure photosynthetic rates of algae by sensing changes in dissolved oxygen concentration (Pandard *et al.*, 1993), or by oxidizing specific carrier compounds that are reduced in the process of electron transport within the algal cell (Rawson *et al.*, 1987). The respiratory uptake of tubificid worms can be monitored by continuous comparisons of dissolved oxygen concentrations before and after the worm-holding cell (Coler *et al.*, 1988). However, most of those methods have not yet been developed for routine use, and the remainder of this discussion pertains to fish.

There are difficulties in all continuous biomonitoring systems. They require sophisticated monitoring and recording equipment. Fish must be fed, tanks must be cleaned, and dead, diseased, or unhealthy organisms must be removed and replaced. The "noise" from swimming or feeding activity can obscure the signals of gill irrigation. There is also the need to program the system for varying activity and response thresholds according to time of day, season, or physiological status. The magnitude of response that is to be considered environmentally significant must be decided. That requires some calibration of response in the monitoring system with toxic effects in standard tests.

The essence of good design in these systems is to find the balance between sensitivity and noise or false alarms (Cairns *et al.*, 1980). This is both a statistical and management problem. Two approaches have been used. (1) A single group of fish is maintained in the wastewater and stress is indicated when the response deviates from an historical baseline, continually updated. (2) A treatment group of fish is maintained in the wastewater while a control group is held in clean source water; an effect is detected by comparing the two groups for a given time-period. In both methods, fish are housed separately to avoid confounding by interactions between individuals.

Statistical analysis is complicated for the single-group system because successive measurements on each fish are not independent. A simple approach is to establish a baseline mean and variance for a time when toxic stress is presumed to be absent, and compare monitoring data with that, using simple parametric statistics [59]. Time-series analysis has been recommended to regularly update the baseline, but unfortunately, statistical analysis of the time-series data is unstandardized and is generally a challenge (Thompson *et al.*, 1982; Westlake *et al.*, 1983; Hayward *et al.*, 1988). The two-group method allows simple, nonparametric comparisons of treatment and control responses It also allows for the background "noise" because variance introduced by feeding or human disturbance is shared by both groups (Hayward *et al.*, 1988). The drawback is that wastewater and control water often differ in temperature, oxygen, or solids, which affect respiration and other functions (Hayward *et al.*, 1988).

Continuous monitoring systems have seen some application in industry [60]. They can also be used for monitoring receiving waters, by diverting surface water through tanks in a mobile laboratory. There is (or was) a consulting company in the USA which specialized in such a mobile system, using behaviour of fish and other organisms to assess effects in effluent streams and surface waters.

---

[59] In fact, little is simple in the statistical treatment. Even the use of variance is difficult, because variability between individual fish tends to increase when they are stressed. Hayward *et al.* (1988) even suggest using the increase in variance as the measurement of stress response, rather than the increase in mean response.

[60] Probably the earliest Canadian example was at Oakville, Ontario. In the 1950s, final effluent at the Cities Service oil refinery passed through an aquarium of fish in the entrance of the office building. The fish were "monitored" by people passing by.

References

Abernethy, S.G. and G.F. Westlake, 1991. Guidelines for pH adjustment of effluent samples for toxicity identification and reduction evaluations. Ontario Ministry of Environment, Water Resources Branch, Toronto, Ont.

Addison, R.F., 1984. p. 52-60, In: *Contaminant effects on fisheries.* V.W. Cairns, P.V. Hodson, and J.O. Nriagu (eds.), John Wiley & Sons, New York, N.Y.

Addison, R.F. and A.J. Edwards, 1988. Hepatic microsomal mono-oxygenase activity in flounder *Platichthys flesus* from polluted sites in Langesundfjord and from mesocosms experimentally dosed with diesel oil and copper. *Mar. Ecol. Prog. Ser., 46:*51–54.

AEP (Alberta Environmental Protection), 1978. Interim Alberta surface water quality guidelines. Alberta Environmental Protection, Edmonton, Alta.

———. 1995. Water quality based effluent limits procedures manual. Alberta Environmental Protection, Environmental Regulatory Service, Edmonton, Alta.

AETE (Aquatic Effects Technology Evaluation), 1997. Toxicity assessment of mining effluents using up-stream or reference site waters and test organism acclimation techniques. Completed for AETE program by ESG International Ltd., Ottawa, Ont., Dec. 1997.

———. 1998a. Summary of cost-effectiveness evaluation of aquatic effects monitoring technologies applied in the 1997 AETE field evaluation program. Completed for AETE by Beak International Inc. and Golder Associates Ltd., Ottawa, Ont., Sept. 1998.

———. 1998b. Addition tool evaluations: Evaluation of receiving water as dilution water in effluent toxicity testing. Completed for AETE program by Beak International Inc., Ottawa, Ont Sept. 1998.

Agriculture Canada, 1987. *The Canadian system of soil classification.* 2nd ed. Pub. No. 1646, Agriculture Canada, Research Branch, Expert Committee on Soil Survey, Ottawa, Ont.

Ahokas, J.T., D.A. Holdway, S.E. Brennan, R.W. Goudey, and H.B. Bibrowska, 1994. MFO activity in carp *(Cyprinus carpio)* exposed to treated pulp and paper mill effluent in Lake Coleman, Victoria, Australia, in relation to AOX, EOX and muscle PCDD/PCDF. *Environ. Toxicol. Chem., 13:*41–50.

Alabaster, J.S. and R. Lloyd (eds.), 1980. *Water quality criteria for freshwater fish.* Butterworths, London and Woburn, Mass.

ALAC/ORSANCO (Aquatic Life Advisory Committee of the Ohio River Valley Water Sanitation Commission), 1955. Aquatic life water quality criteria. First progress report. *Sewage Indust. Wastes, 27*:321–331.

Alden, R.W. III, 1992. Uncertainty and sediment quality assessments: I. Confidence limits for the triad. *Environ. Toxicol. Chem., 11*:637–644.

Aldenberg, T. and W. Slob, 1991. Confidence limits for hazardous concentrations based on logistically distributed NOEC data. RIVM Rept nr. 719102002. Rijksinstituut voor Volksgezondheid en Milieuhygiene [National Inst. of Public Health and Environmental Protection], Bilthoven, The Netherlands.

Alderdice, D.F. and J.R. Brett, 1957. Some effects of kraft mill effluent on young Pacific salmon. *J. Fish. Res. Board Can., 14:*783–795.

Ames, B.N., 1996. The causes of aging and cancer: The misinterpretation of animal cancer tests. *Human Ecol. Risk Assess., 2:* 6–9.

Ames, B.N., J. McCann, and E. Yamaski, 1975. Methods for detecting carcinogens and mutagens with *Salmonella*/mammalian microsome mutagenicity test. *Mutation Res., 31:* 347–364.

Anderson, P.D. and P.A. Spear, 1980. Copper pharmacokinetics in fish gills. II. Body size relationships for accumulation and tolerance. *Water Res., 14:*1107–1111.

Anderson, S.L. and T.J. Norberg-King, 1991. Letter to the editor. Precision of short-term chronic toxicity tests in the real world. *Environ. Toxicol. Chem., 10:*143–145.

Anderson, S.L. and G.C. Wild, 1994. Linking genotoxic responses and reproductive success in ecotoxicology. *Environ. Health Perspect, 102 (Suppl. 12):*9–12.

Anderson, S., W. Sadinski, L. Shugart, P. Brussard, M. Depledge, T. Ford, J. Hose, J. Stegeman, W. Suk, I. Wirgin, and G. Wogan, 1994. Genetic and molecular ecotoxicology: a research framework. *Environ. Health Perspect., 102 (Suppl. 12):*3–8.

Anderson, S., N. Belfiore, and F. Harrison, 1998. Genotoxic effects in aquatic organisms: Methodologies, field-related applications, and the role of genetic diversity. Rept to Atomic Energy Control Board and Environment Canada, August 1998.

Andersson, T., L. Forlin, J. Hardig, and A. Larsson, 1988. Physiological disturbances in fish living in coastal water polluted with bleached kraft pulp mill effluent. *Can. J. Fish. Aquat. Sci., 45:*1525–1536.

Ankley, G.T., 1996. Evaluation of metal/acid-volatile sulfide relationships in the prediction of metal bioaccumulation by benthic macroinvertebrates. *Environ. Toxicol. Chem., 15:*2138–2146.

Ankley, G.T. and L.P. Burkhard, 1992. Identification of surfactants as toxicants in a primary effluent. *Environ. Toxicol. Chem., 11:*1235–1248.

Ankley, G.T. and R.E. Reinert, 1988. Effects of commercial versus synthetic diets on hepatic xenobiotic-metabolizing enzymes in channel catfish. *Mar. Environ. Res., 24:*41–44.

Ankley, G.T. and M.K. Schubauer-Berigan, 1995. Background and overview of current sediment toxicity identification evaluation procedures. *J. Aquat. Ecosystem Health, 4:*133–149.

Ankley, G.T., R.A. Hoke, J.P. Giesy, and P.V. Winger, 1989. Evaluation of the toxicity of marine sediments and dredge spoils with the Microtox® bioassay. *Chemosphere , 18:*2069–2075.

Ankley, G.T., A. Katko, and J.W. Arthur, 1990. Identification of ammonia as an important sediment-associated toxicant in the lower Fox River and Green Bay, Wisconsin. *Environ. Toxicol. Chem., 9:*313–322.

Ankley, G.T., G.L. Phipps, E.N. Leonard, D.A. Benoit, V.R. Mattson, P.A. Kosian, A.M. Cotter, J.R. Dierkes, D.J. Hansen, and J.D. Mahony, 1991a. Acid-volatile sulphide as a factor mediating cadmium and nickel bioavailability in contaminated sediments. *Environ. Toxicol. Chem., 10:*1299–1308.

Ankley, G.T., M.K. Schubauer-Berigan, and J.R. Dierkes, 1991b. Predicting the toxicity of bulk sediments to aquatic organisms with aqueous test fractions: pore water vs. elutriate. *Environ. Toxicol. Chem., 10:*1359–1366.

Ankley, G.T., D.M. Di Toro, D.J. Hansen, and W.J. Berry, 1996. Assessing the ecological risk of metals in sediments. *Environ. Toxicol. Chem., 15:*2053–2055.

Ankley, G., E. Mihaich, R. Stahl, D. Tillitt, T. Colborn, S. McMaster, R. Miller, J. Bantle, P. Campbell, N. Denslow, R. Dickerson, L. Folmar, M. Fry, J. Giesy, L. E. Gray, P. Guiney, T. Hutchinson, S. Kennedy, V. Kramer, G. LeBlanc, M. Mayes, A. Nimrod, R. Patino, R. Peterson, R. Purdy, R. Ringer, P. Thomas, L. Touart, G. Van Der Kraak, and T. Zacharewski, 1998. Overview of a workshop on screening methods for detecting potential (anti-) estrogenic/androgenic chemicals in wildlife. *Environ. Toxicol. Chem., 17:*68–87.

APHA, AWWA, WEF (Amer. Public Health Assoc., Amer. Water Works Assoc., and Water Environment Federation), 1995. Standard methods for the examination of water and wastewater. 19th ed., APHA, Washington, D.C.

API (American Petroleum Institute), 1988. Fathead minnow 7-day test: Round robin study. Intra- and interlaboratory study to determine the reproducibility of the 7-day fathead minnow larval survival and growth test. API Pub. No. 4468. Amer. Petroleum Inst., Health & Environ. Sci. Dept, Washington, D.C.

Arcand-Hoy, L.D. and W.H. Benson, 1998. Fish reproduction: An ecologically relevant indicator of endocrine disruption. *Environ. Toxicol. Chem., 17:*49–57.

Armitage, P.D., R.J.M. Gunn, M.T. Furse, J.F. Wright, and D. Moss, 1987. The use of prediction to assess macroinvertebrate response to river regulation. *Hydrobiologia, 144:*25–32.

ASTM (American Society for Testing and Materials), 1997a. Annual book of ASTM standards. Section 11, Water and environmental technology, vol. 11.5: Biological effects and environmental fate; biotechnology; pesticides. ASTM, Philadelphia, Pa.

———. 1997b. Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphibians. Annual Book of Standards, Section II, Water and Environmental Technology, Volume 11.05, ASTM Designation E 729-96, p. 220–240, ASTM, Philadelphia, Pa.

Atema, J. and L.S. Stein, 1974. Effects of crude oil in the feeding behaviour of the lobster, *Homarus americanus. Environ. Pollut., 6:*77–88.

ATSDR (Agency for Toxic Substances and Disease Registry), 1998. CRCnetBASE™ 1999. CRC Press LLC, Boca Raton, Fla.

Ausmus, B.S. and E.G. O'Neill, 1978. Comparison of carbon dynamics of three microcosm substrates. *Soil Biol. Biochem., 10:*425–430.

Ausley, L.W., 1996. Effluent toxicity testing variability. p. 157–171, In: *Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins (eds.). Soc. Environ. Toxicol. Chem., SETAC Spec. Pub. Series, SETAC Press, Pensacola, Fla.

Australia, 1989. Pulp and paper industry package. Commonwealth of Australia, Departments of Industry, Technology and Commerce; Arts, Sport, the Environment, Tourism and Territories; Primary Industries and Energy; and Commonwealth Scientific and Industrial Research Organization. Canberra, Australia, 76 p.

Bailey, H.C., J.L. Miller, M.J. Miller, and B.S. Dhaliwal, 1995. Application of toxicity identification procedures to the echinoderm fertilization assay to identify toxicity in a municipal effluent. *Environ. Toxicol. Chem., 14:*2181–2186.

Bailey, H.C., J.L. Miller, M.J. Miller, L.C. Wiborg, L. Deanovic, and T. Shed, 1997. Joint acute toxicity of Diazinon and Chlorpyrifos to *Ceriodaphnia dubia. Environ. Toxicol. Chem., 16:*2304–2308.

Baird, R.B., R. Berger, and J. Gully, 1996. Improvements in point estimation methods and application to controlling aquatic toxicity test reliability. p. 103–130, In: *Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins (eds.). Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Banas, W.P. and J.B. Sprague, 1986. Absence of acclimation to Parathion by rainbow trout during sublethal exposure. *Water Res., 20:*1229–1232.

Barber, T.R., D.J. Chappie, D.J. Duda, P.C. Fuchsman, and B.L. Finley, 1998. Using a spiked sediment bioassay to establish a no-effect concentration for dioxin exposure to the amphipod *Ampelisca abdita. Environ. Toxicol. Chem., 17:*420–424.

Barbour, M.T., J.M. Diamond, and C.O. Yoder, 1996. Biological assessment strategies: applications and limitations. p. 245–270, In: *Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts.* D.R Grothe, K.L. Dickson, and D.K. Reed-Judkins (eds.). Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Barnthouse, L.W., G.W. Suter II, and A.E. Rosen, 1990. Risks of toxic contaminants to exploited fish populations: influence of life history, data uncertainty and exploitation intensity. *Environ. Toxicol. Chem., 9:*297–311.

Baud-Grasset, F., S. Baud-Grasset, and S.I. Safferman, 1993. Evaluation of the bioremediation of a contaminated soil with phytotoxicity tests. *Chemosphere, 26:*1365–1374.

B.C. MELP (British Columbia Ministry of Environment, Lands and Parks), 1998. British Columbia water quality guidelines (criteria): 1998 edition. B.C. MELP, Water Management Branch, Water Quality Section, Victoria, B.C. 30 p.

Beak (Beak Consultants Ltd.), 1979. Toxicity of effluents from sulphite pulping operations practicing recovery and biological treatment. Economic and Tech. Rev. Rept No. EPS 3-WP-79-9, Environment Canada, Environ. Prot. Service, Water Pollut. Control Dir., Ottawa, Ont. 51 p.

Beanlands, G.E. and P.N. Duinker, 1983. An ecological framework for environmental impact assessment in Canada. Inst. Resource and Environ. Studies, Dalhousie Univ., Halifax, and Canadian Federal Environmental Review Office, Hull, Quebec. 132 p.

Becker, D.S., G.R. Bilyard, and T.C. Ginn, 1990. Comparisons between sediment bioassays and alterations of benthic macroinvertebrate assemblages at a marine superfund site: Commencement Bay, Washington. *Environ. Toxicol. Chem., 9:*669–685.

Bedard, D., A. Haydon, and D. Persaud, 1992. Ontario Ministry of the Environment laboratory sediment biological testing protocol. PIBS 2067E, Log 92-2309-074, OMOE, Water Resources Br., Rexdale, Ont., 26 p.

BEIR (Biological Effects of Ionizing Radiation), 1973. Biological effects of ionizing radiation report. *Bull. Atomic Scientists, March:*47–49.

Belfroid, A., W. Seinen, K. van Gestel, and J. Hermens, 1993. The acute toxicity of chlorobenzenes for earthworms *(Eisenia andrei)* in different exposure systems. *Chemosphere, 26:*2265–2277.

Benoit, D.A., P.K. Sibley, J.L. Juenemann, and G.T. Ankley, 1997. *Chironomus tentans* life-cycle test: design and evaluation for use in assessing toxicity of contaminated sediments. *Environ. Toxicol. Chem., 16:*1165–1176.

Benson, W.H. and R.T. Di Giullo, 1992. Biomarkers in hazard assessment of contaminated sediments. p. 241–265, In: *Sediment toxicity assessment.* G.A. Burton, Jr. (ed.), Lewis Pub., Boca Raton, Fla.

Benton, M.J., M.L. Malott, S.S. Knight, C.M. Cooper, and W.H. Benson, 1995. Influence of sediment composition on apparent toxicity in a solid-phase test using bioluminescent bacteria. *Environ. Toxicol. Chem., 14:* 411–414.

Bergman, H.L. and E.J. Dorward-King, 1997. Reassessment of metals criteria for aquatic life protection. (Proc. of the Pellston Workshop) Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Bermingham, N., C. Blaise, and G. Costan, 1995. Use of microscale aquatic toxicity tests to set "green standards" for general purpose cleaners. *Proc. 22nd Ann. Aquatic Toxicity Workshop*, St. Andrews, N.B., Oct. 1-5, 1995.

Berry, W.J., D.J. Hansen, J.D. Mahony, D.L. Robson, D.M. Di Toro, B.P. Shipley, B. Rogers, J.M. Corbin, and W.S. Boothman, 1996. Predicting the toxicity of metal-spiked laboratory sediments using acid-volatile sulfide and interstitial water normalizations. *Environ. Toxicol. Chem., 15:*2067–2079.

Bertelsen, S.L., A.D. Hoffman, C.A. Gallinat, C.M. Elonen, and J.W. Nichols, 1998. Evaluation of log K_{OW} and tissue lipid content as predictors of chemical partitioning to fish tissues. *Environ. Toxicol. Chem., 17:* 1447–1455.

Besser, J.M., C.G. Ingersoll, E.N. Leonard, and D.R. Mount, 1998. Effect of zeolite on toxicity of ammonia in freshwater sediments: implication for toxicity identification evaluation procedures. *Environ. Toxicol. Chem., 17:* 2310–2317.

Birge, W.J., J.A. Black, T.M. Short, and A.G. Westerman, 1989. A comparative ecological and toxicological investigation of a secondary wastewater treatment plant effluent and its receiving stream. *Environ. Toxicol. Chem. 8:* 437–450.

Birtwell, I.K., 1989. Comments on sensitivity of salmonids to reduced levels of dissolved oxygen and to pulp mill pollution in Neroutsos Inlet, British Columbia. *Canadian Tech. Rept Fish. Aquat Sci.,* 1695, 27 p.

Bitton, G. and B.J. Dutka, 1986. *Toxicity testing using microorganisms. Vol. 1.* CRC Press, Boca Raton, Fla.

BKH (Bongaerts, Kuyper, & Huiswaard Consulting) Engineers, 1990. Ecotoxicological effects assessment: Extrapolation from single species toxicity data. Prepared for the Netherlands Ministry of Housing, Physical Planning and Environment, The Hague, The Netherlands. 97 p.

Black, J., 1983. Field and laboratory studies of environmental carcinogenesis in Niagara River fish. *J. Great Lakes Res. 9:*326–334.

Blaise, C., R. Legault, N., Bermingham, R. van Coillie, and P. Vasseur, 1984. Microtest mesurant l'inhibition de la croissance des algues (CI50) par le dosage de l'ATP. *Sci. Tech. Eau, 17:*245–250.

Blaise, C., N. Bermingham, and R. Van Coillie, 1985. The integrated ecotoxicological approach to assessment of ecotoxicity. *Water Quality Bull., 10 (1):*3–10, 60–61.

Blaise, C., G. Sergy, P. Wells, N. Bermingham, and R. Van Coillie, 1988. Biological testing–development, application, and trends in Canadian Environmental Protection laboratories. *Toxicity Assess., 8:* 385–406.

Blaise, C., J.-F. Férard, and P. Vasseur, 1998. Microplate toxicity tests with microalgae: A review. p. 269–288, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* Wells, P.G., K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Blenkinsopp, S., P. Boileau, D. Kyle, G. Sergy, and M. Fingas, 1996. How to prepare water accommodated fractions from petroleum hydrocarbons for use in aquatic toxicity testing -- the basics. *Proc. of 19th AMOP Meeting,* June 12-14, 1996, Vol. 1:515–528, Calgary, Alberta.

Bol, J., H.J.M. Verhaar, C.J. van Leeuwen, and J.L.M. Hermens, 1993a. Predictions of the aquatic toxicity of high-production-volume-chemicals. Part A: Introduction and methodology. The Netherlands Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, Diretoraat-Generaal Milieubeheer, nr. 1993/9a, 11 p. + encl.

———. 1993b. Predictions of the aquatic toxicity of high-production-volume-chemicals. Part B: prediction. The Netherlands Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, Diretoraat-Generaal Milieubeheer, nr. 1993/9b, 125 p.

Bonsor, N., N. McCubbin, and J.B. Sprague, 1988. Kraft mill effluents in Ontario. Prepared for Ontario Ministry of Environment, Pulp and Paper Sector of MISA, Technical Advisory Committee, Toronto, Ont.

Borgmann, U., R. Cove, and C. Loveridge, 1980. Effect of metals on the biomass production kinetics of freshwater copepods. *Can. J. Fish. Aquat. Sci., 37:*567–575.

Borgmann, U. W.P. Norwood, and I.M. Babirad, 1991. Relationship between chronic toxicity and bioaccumulation of cadmium in *Hyalella azteca*. *Can. J. Fish. Aquat. Sci., 48:*1055–1060.

Botosaneanu, L. (ed.), 1986. Stygofauna mundi. A faunistic, distributional, and ecological synthesis of the world fauna inhabiting subterranean waters (including the marine interstitial). E.J. Brill/Dr. W. Backhuys, Leiden, the Netherlands.

Bradley, R.W. and J.B. Sprague, 1985. The influence of pH, water hardness, and alkalinity on the acute lethality of zinc to rainbow trout *(Salmo gairdneri)*. *Can. J. Fish. Aquat. Sci., 42:* 731–736.

Brady, N.C., 1974. *The nature and properties of soils*. 8th ed. MacMillan Pub. Co., New York, N.Y. 639 p.

Bretthauer, E.W., 1992. The challenge of ecological risk assessment. *Environ. Toxicol. Chem., 11:*1661–1662.

Broderius, S.J., 1991. Modeling the joint toxicity of xenobiotics to aquatic organisms: basic concepts and approaches. p. 107–127, In: Aquatic toxicology and risk assessment: 14th vol. M.A. Mayes and M.G. Barron (eds.), ASTM STP 1124, Amer. Soc. Testing and Materials, Philadelphia, Pa.

Bruce, R.D. and D.J. Versteeg, 1992. A statistical procedure for modeling continuous toxicity data. *Environ. Toxicol. Chem., 11:*1485–1494.

Buikema, A.L. Jr. and E.F. Benfield, 1979. Use of macroinvertebrate life history information in toxicity tests. *J. Fish. Res. Board Can., 36:*321–328.

Buikema, A.L. Jr. and J.R. Voshell Jr., 1993. Toxicity studies using freshwater benthic macroinvertebrates. p. 344–398, In: *Freshwater biomonitoring and benthic macroinvertebrates.* D.M. Rosenberg and V.H. Resh (eds.), Chapman & Hall, New York, N.Y.

Buikema, A.L. Jr., B.R. Niederlehner, and J. Cairns Jr., 1982. Biological monitoring, Part IV – toxicity testing. *Water Res., 16:*239–262.

Burmaster, D.E., C.A. Menzie, and J.S. Freshman, 1991. Assessment of methods for estimating aquatic hazards at superfund-type sites: a cautionary tale. *Environ. Toxicol. Chem., 10:*827–842.

Burns, L.A., C.G. Ingersoll, and G.A. Pascoe, 1994. Ecological risk assessment: Application of new approaches and uncertainty analysis. *Environ. Toxicol. Chem., 13:*1873–1874.

Burton, G.A. Jr., 1991. Assessing the toxicity of freshwater sediments. *Environ. Toxicol. Chem., 10:*1587–1628.

———. (ed.), 1992. *Sediment toxicity assessment.* Lewis Pub., Boca Raton, Fla.

Burton, G.A. Jr., T.J. Norberg-King, C.G. Ingersoll, G.T. Ankley, P.V. Winger, J. Kubitz, J.M. Lazorchak, M.E. Smith, E. Greer, F.J. Dwyer, D.J. Call, K.E. Day, P. Kennedy, and M. Stinson, 1996. Interlaboratory study of precision: *Hyalella azteca* and *Chironomus tentans* freshwater sediment toxicity assays. *Environ. Toxicol. Chem., 15:*1335–1342.

Cairns, J. Jr., 1983. Are single species toxicity tests alone adequate for estimating environmental hazard? *Hydrobiologia, 100:*47–57.

———. (ed.), 1985. *Multispecies toxicity testing.* Pergamon Press, New York, N.Y.

———. 1986. What is meant by validation of predictions based on laboratory toxicity tests? *Hydrobiologia, 137:* 271–278.

———. 1988a. What constitutes field validation of predictions based on laboratory evidence? p. 361–368, In: *Aquatic toxicology and hazard assessment: 10th vol.* W.J. Adams, G.A. Chapman, and W.G. Landis (eds.), Amer. Soc. Testing and Materials, Philadelphia, Pa., ASTM STP 971.

———. 1992. The threshold problem in ecotoxicology. *Ecotoxicology, 1:*3-16.

Cairns, J. Jr. and P.V. McCormick, 1992. Developing an ecosystem-based capability for ecological risk assessments. *The Environ. Professional, 14:*186–196.

Cairns, J., K.L. Dickson, and A.W. Maki, 1978. Summary and conclusions. p. 191–197, In: *Estimating the hazard of chemical substances to aquatic life.* J. Cairns, K.L. Dickson, and A.W. Maki (eds.). Amer. Soc. Testing and Materials, Philadelphia, Pa., ASTM STP 657.

Cairns, J. Jr., K.W. Thompson, J.D. Landers Jr., M.T. McKee, and A.C. Hendricks, 1980. Suitability of some freshwater and marine fishes for use with a minicomputer interfaced biological monitoring system. *Water Resour. Bull., 16:*421–427.

Cairns, J. Jr., B.R. Niederlehner, and E.P. Smith, 1992. The emergence of functional attributes as endpoints in ecotoxicology. p. 111–128, In: *Sediment toxicity assessment*. G.A. Burton, Jr. (ed.), Lewis Pub., Boca Raton, Fla.

Cairns, J. Jr., B.R. Niederlehner, and E.P. Smith, 1998. Correspondence of a microscale toxicity test to responses to toxicants in natural systems. p. 539–547, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Calabrese, E.J. and L.A. Baldwin, 1993. *Performing ecological risk assessments.* Lewis Pub., Ann Arbor, Mich., 288 p.

California, 1972. Water quality control plan for ocean waters of California. State of California, The Resources Agency, State Water Resources Control Board, Sacramento, California. Unnumbered rept.

Calleja, M.C. and G. Persoone, 1993. The influence of solvents on the acute toxicity of some lipophilic chemicals to aquatic invertebrates. *Chemosphere, 26:*2007–2022.

Campbell, P.G.C. and P.M. Stokes, 1985. Acidification and toxicity of metals to aquatic biota. *Can. J. Fish. Aquat. Sci., 42:*2034–2049.

Campbell, P.M., and T.H. Hutchinson, 1998. Wildlife and endocrine disrupters: requirements for hazard identification. *Environ. Toxicol. Chem., 17:*127–135.

Canfield, T.J., N.E. Kemble, W.G. Brumbaugh, F.J. Dwyer, C.G. Ingersoll, and J.F. Fairchild, 1994. Use of benthic invertebrate community structure and the sediment quality triad to evaluate metal contaminated sediment in the Upper Clark Fork River, Montana. *Environ. Toxicol. Chem., 13:*199–2012.

Canton, S.P., and W.D. Van Derveer, 1997. Selenium toxicity to aquatic life: an argument for sediment-based water quality criteria. *Environ. Toxicol. Chem., 16:*1255–1259.

Carder, J.P., and K.D. Hoagland, 1998. Combined effects of Alachlor and Atrazine on benthic algal communities in artificial streams. *Environ. Toxicol. Chem. 17:*1415–1420.

Carlson, R.W. and R.A. Drummond, 1978. Fish cough response – a method for evaluating quality of treated complex effluents. *Water Res., 12:*1–6.

Carlson, A.R., G.L. Phipps, V.R. Mattson, P.A. Kosian, and A.M. Cotter, 1991. The role of acid-volatile sulphide in determining cadmium bioavailability and toxicity in freshwater sediments. *Environ. Toxicol. Chem., 10:*1309–1320.

Carr, R.S., 1998. Marine and estuarine porewater toxicity testing. p. 5232–5238, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Carr, R.S., E.R. Long, H.I. Windom, D.C. Chapman, G. Thursby, G.M. Sloane, and D.A. Wolfe, 1996. Sediment quality assessment studies of Tampa Bay, Florida. *Environ. Toxicol. Chem., 15:*1218–1231.

Carrier, R. and T.L. Beitinger, 1988. Reduction in thermal tolerance of *Notropis lutrensis* and *Pimephales promelas* exposed to cadmium. *Water Res., 22:*511–515.

CCME (Canadian Council of Ministers of the Environment), 1991a. Interim Canadian environmental quality criteria for contaminated sites. Rept CCME EPC-CS34, CCME, Subcommittee on Environmental Quality Criteria for Contaminated Sites, Winnipeg, Man. 20 p.

———. 1991b. Appendix IX. A protocol for the derivation of water quality guidelines for the protection of aquatic life. (April 1991). In: Canadian water quality guidelines, Canadian Council of Resource and Environment Ministers. 1987. Prepared by the Task Force on Water Quality Guidelines. [Reprinted in CCME 1998].

———. 1992. National classification system for contaminated sites. Rept no. CCME EPC-CS39E. Prepared by the CCME Subcommittee on Classification of Contaminated Sites, CCME, Winnipeg, Man.,54 p.

———. 1993a. Appendix XV. Canadian water quality guidelines. Updates (October 1993). Protocols for deriving water quality guidelines for the protection of agricultural water uses. In: Canadian water quality guidelines. CCREM. 1987. Prepared by the Task Force on Water Quality Guidelines. 18 p.

———. 1993b. Guidance manual on sampling, analysis, and data management for contaminated sites. Volume I: main report. Rept no. CCME-EPC-NCS-62E, CCME, Winnipeg, Man., 71 p.

———. 1993c. Guidance manual on sampling, analysis, and data management for contaminated sites. Volume II: analytical method summaries. Rept no. CCME-EPC-NCS-66E. CCME, Winnipeg, Man.,171 p.

———. 1995. Protocol for the derivation of Canadian sediment quality guidelines for the protection of aquatic life.

Rept CCME EPC-98E. CCME, Task Group on Water Quality Guidelines, Ottawa, Ont. 38 p.

———. 1996a. Guidance manual for developing site-specific soil quality remediation objectives for contaminated sites in Canada. Unnumbered rept., CCME, Subcommittee on Environmental Quality Criteria for Contaminated Sites. Winnipeg, Man. 50 p.

———. 1996b. A protocol for the derivation of environmental and human health soil quality guidelines. Unnumbered rept., CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, Winnipeg, Man. 169 p.

———. 1996c. Canadian environmental quality guidelines for dioxins and furans. CCME Task Group on Water Quality Guidelines, Ottawa. Final draft.

———. 1996d. A framework for ecological risk assessment. General guidance. Unnumbered rept, CCME, Subcommittee on Environmental Quality Criteria for Contaminated Sites, Winnipeg, Man. 50 p.

———. 1996e. A framework for ecological risk assessment. Technical appendices. Unnumbered rept., CCME, Subcommittee on Environmental Quality Criteria for Contaminated Sites, Winnipeg, Man. 140 p.

———. 1997a. Recommended Canadian soil quality guidelines. Unnumbered rept., CCME, Subcommittee on Environmental Quality Criteria for Contaminated Sites. Winnipeg, Man. 185 p.

———. 1997b. Protocol for the derivation of Canadian tissue residue guidelines for the protection of wildlife that consume aquatic biota. CCME Task Group on Water Quality Guidelines, Ottawa.

———. 1998. Canadian environmental quality guidelines, Canadian Council of Ministers of the Environment, Winnipeg, Man.

CCREM (Canadian Council of Resource and Environment Ministers), 1987. Canadian water quality guidelines. Prepared by the Task Group on Water Quality Guidelines, Ottawa, Ont.

CEPA (Canadian Environmental Protection Act), 1988. Interim guidance for the application of rapidly rendered harmless (RRH) and trace contaminants. Canadian Environmental Protection Act, R.S., c. 16 (4th Supp.), Subsection 71(3).

Chapman, G.A., B.S. Anderson, A.J. Bailer, R.B. Baird, R. Berger, D.T. Burton, D.L. Denton, W.L. Goodfellow Jr., M.A. Heber, L.L. McDonald, T.J. Norberg-King, P.J. Ruffier, 1996. Methods and appropriate endpoints, discussion synopsis. p. 51–82, In: *Whole effluent toxicity testing: an evaluation of methods and prediction of receiving system impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins (eds.). Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Chapman, P.M, 1986. Sediment quality criteria from the sediment quality triad: An example. *Environ. Toxicol. Chem., 5:*957–964.

———. 1989a. A bioassay by any other name might not smell the same. *Environ. Toxicol. Chem., 8:*557.

———. 1989b. Review. Current approaches to developing sediment quality criteria. *Environ. Toxicol. Chem., 8:* 589–599.

———. 1990. The sediment quality triad approach to determining pollution-induced degradation. *Science Total Environ., 97/98:*815–825.

———. 1995. Ecotoxicology and pollution -- Key issues. *Marine Pollution Bull., 31:*167–177.

———. 1996. Presentation and interpretation of sediment quality triad data. *Ecotoxicology, 5:*327–339.

Chapman, P.M. and R. Fink, 1984. Effects of Puget Sound sediments and their elutriates on the life cycle of *Capitella capitata. Bull. Environ. Contam. Toxicol., 33:*451–459.

Chapman, P.M., R.N. Dexter, and E.R. Long, 1987. Synoptic measures of sediment contamination, toxicity and infaunal community structure (the sediment quality triad) in San Francisco Bay. *Mar. Ecol. Prog. Ser., 37:*75–96.

Chapman, P.M., E.A. Power, R.N. Dexter, and H.B. Anderson, 1991. Evaluation of effects associated with an oil platform, using the sediment quality triad. *Environ. Toxicol. Chem., 10:*407–424.

Chapman, P.M., E.A. Power, and G.A. Burton Jr., 1992. Integrative assessments in aquatic ecosystems. p. 313–340, In: *1992. Sediment toxicity assessment*. G.A. Burton, Jr. (ed.), Lewis Pub., Boca Raton, Fla.

Chapman, P.M., A. Fairbrother, and D. Brown, 1998. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ. Toxicol. Chem., 17:*99–108.

Chevrier, A. and L. Porebski, 1993. Development of marine bioassays for use in pollution prevention – An example: Ocean disposal. *Symp. on Environ. Effects Monitoring (EEM)*, Quebec (Quebec), October 1993.

Chevrier, A. and P. Topping, 1998. National guidelines for monitoring dredged and excavated material at ocean disposal sites. Unnumbered rept, Dec. 1998. Environment Canada, Marine Environ. Div., Disposal at Sea Program, Ottawa.

Christensen, E.R. and C-Y Chen, 1991. Modeling of combined toxic effects of chemicals. *Toxic Substances J., 11:*1–63.

Clements, W.H. and P.M. Kiffney, 1996. Validation of whole effluent toxicity tests: Integrated studies using field assessments, microcosms, and mesocosms. p. 229–244, In: Whole effluent toxicity testing: an evaluation of methods and prediction of receiving system impacts. D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins (eds). Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Clements, W.H., D.S. Cherry, and J. Cairns Jr., 1988. Structural alterations in aquatic insect communities exposed to copper in laboratory streams. *Environ. Toxicol. Chem., 7:*715–722.

Coler, R.A., M.S. Coler, and P.T. Kostecki, 1988. Tubificid behaviour as a stress indicator. *Water Res., 22:*263–267.

Costan, G., N. Bermingham, C. Blaise, and J.F. Ferard, 1993. Potential Ecotoxic Effects Probe (PEEP): A novel index to assess and compare the toxic potential of industrial effluents. *Environ. Toxicol. Water Qual., 8:*115–140.

Cripe, G.M., A. Ingley-Guezou, L.R. Goodman, and J. Forester, 1989. Effect of food availability on the acute toxicity of four chemicals to *Mysidopsis bahia* (Mysidacea) in static exposures. *Environ. Toxicol. Chem., 8:*333–338.

Crossland, N.O. and G.C. Mitchell, 1992. Use of outdoor artificial streams to determine threshold toxicity concentrations for a petrochemical effluent. *Environ. Toxicol. Chem., 11:*49–59.

Curtis, G.L, 1990. Recovery of offshore lake trout *(Salvelinus namaycush)* population in eastern Lake Superior. *J. Great Lakes Res., 16:*7–15.

Cusimano, R.F., D.F. Brakke, and G.A. Chapman, 1986. Effects of pH on the toxicities of cadmium, copper, and zinc to steelhead trout *(Salmo gairdneri). Can. J. Fish. Aquat. Sci., 43:*1497–1503.

Danielopol, D.L., 1992. New perspective in ecological research of groundwater organisms. p. 15–22, In: Proc. First International Conference on Ground Water Ecology, April 26-29, 1992, J.A. Stanford and J.J. Simons (eds.), Tampa, Fla.

Dankwardt, A., S. Pullen, and B. Hock, 1998. Immunoassays: applications for the aquatic environment. p. 13–29, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Dasappa, S.M. and R. Loehr, 1991. Toxicity reduction in contaminated soil bioremediation processes. *Water Res., 25:*1121–1130.

Davies, P.H. and J.D. Woodling, 1980. Importance of laboratory-derived metal toxicity results in predicting in-stream response of resident salmonids. p. 281–299, In: *Aquatic toxicology.* ASTM STP 707. J.G. Eaton, P.R. Parrish, and A.C. Hendricks (eds.), ASTM, Philadelphia Pa.

Day, K.E., 1991. Effects of dissolved organic carbon on accumulation and acute toxicity of fenvalerate, deltamethrin and cyhalothrin to *Daphnia magna* (Straus). *Environ. Toxicol. Chem., 10:*91–101.

Day, K., T. Reynoldson, and D. Rosenberg, 1996. The reference condition: Implications for biomonitoring of water and sediment quality in Canada. Soc. Environ. Toxicol. Chemistry, Pensacola, Fla., *SETAC News, Nov. 1996:*15–16.

Denizeau, F., 1998. The use of fish cells in the toxicological evaluation of environmental contaminants. p. 113–128, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G., Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

de Peyster, A. and K. Day, 1998. Ecological risk assessment: A meeting of policy and science. Soc. Environ. Toxicol. Chem., SETAC Press, Pensacola, Fla., 225 p.

Detenbeck, N.E., P.W. DeVore, G.J. Niemi, and A. Lima, 1992. Recovery of temperate-stream fish communities from disturbance: A review of case studies and synthesis of theory. *Environ. Manage., 16:*33–53.

Dewey, S.L., 1986. Effects of the herbicide atrazine on aquatic insect community structure and emergence. *Ecology, 67:*148–162.

DeWitt, T.H., G.R. Ditsworth, and R.C. Swartz, 1988. Effects of natural sediment features on survival of the phoxocephalid amphipod *Rheopoxynius abronius*. *Mar. Environ. Res., 25:*99–124.

Dewitt, T.H., R.C. Swartz, and J.0. Lamberson, 1989. Measuring the acute toxicity of estuarine sediments. *Environ. Toxicol. Chem. 8:*1035–1048.

DFO (Department of Fisheries and Oceans), 1992. Fisheries Act. Pulp and paper effluent regulations. Canada Gazette Part II, Vol. 126, No. 11, May 7, 1992, p. 1967–2006.

Diaz, R.J., 1992. Ecosystem assessment using estuarine and marine benthic community structure. p. 67–85, In: *Sediment toxicity assessment.* G.A. Burton (ed.), Lewis Publishers, Boca Raton, Fla.

Dickson, K.L., W.T. Waller, J.H. Kennedy, and L.P. Ammann, 1992. Assessing the relationship between ambient toxicity and instream biological response. *Environ. Toxicol. Chem., 11:*1307–1322.

Di Toro, D.M., J.D. Mahony, D.J. Hansen, K.J. Scott, M.B. Hicks, S.M. Mayr, and M.S. Redmond, 1990. Toxicity of cadmium in sediments: The role of acid volatile sulfide. *Environ. Toxicol. Chem., 9:*1487–1502.

Di Toro, D.M., C.S. Zarba, D.J. Hansen, W.J. Berry, R.C. Swartz, C.E. Cowan, S.P. Pavlou, H.E. Allen, N.A. Thomas, and P.R. Paquin, 1991. Technical basis for establishing sediment quality criteria for nonionic organic chemicals by using equilibrium partitioning. *Environ. Toxicol. Chem., 10:*1541–1586.

Dixon, D.G. and J.W. Hilton, 1981. Influence of available dietary carbohydrate content on tolerance of waterborne copper by rainbow trout. *J. Fish. Biol., 19:*509–517.

———. 1985. Effects of available dietary carbohydrate and water temperature on the chronic toxicity of waterborne copper to rainbow trout *(Salmo gairdneri). Can. J. Fish. Aquat. Sci., 42:*1007–1013.

Dixon, D.G. and J.B. Sprague, 1981a. Acclimation to copper by rainbow trout *(Salmo gairdneri)* – A modifying factor in toxicity. *Can. J. Fish. Aquat. Sci., 38:*880–888.

———. 1981b. Copper bioaccumulation and hepatoprotein synthesis during acclimation to copper by juvenile rainbow trout. *Aquatic Toxicol., 1:*69–81.

Dixon, W.J. and F.J. Massey Jr., 1969. *Introduction to statistical analysis.* 3rd ed., McGraw-Hill, New York.

Doe, K.G., W.R. Ernst, W.R. Parker, G.R.J. Julien, and P.A. Hennigar, 1988. Influence of pH on the acute lethality of fenitrothion, 2,4-D and aminocarb and some pH-altered sublethal effects of aminocarb on rainbow trout *(Salmo gairdneri). Can. J. Fish. Aquat. Sci., 45:*287–293.

Douglas, W.S. and M.T. Horne, 1997. The interactive effects of essential ions and salinity on the survival of *Mysidopsis bahia* in 96-h acute toxicity tests of effluents discharged to marine and estuarine receiving waters. *Environ. Toxicol. Chem., 16:*1996–2001.

Donnelly, K.C., K.W. Brown, C.S. Anderson, and J.C. Thomas, 1991. Bacterial mutagenicity and acute toxicity of solvent and aqueous extracts of soil samples from an abandoned chemical manufacturing site. *Environ. Toxicol. Chem., 10:*1123–1131.

Dowdy, D.L. and T.E. McKone, 1997. Predicting plant uptake of organic chemicals from soil or air using octanol/water and octanol/air partition ratios and a molecular connectivity index. *Environ. Toxicol. Chem., 16:*2448–2456.

Durand, M.J., V. Cruciani, S. Alexandre, J.C. Hoflack, H. Bessi, C. Blaise, and P. Vasseur, 1998. *In vitro* assays for detection of nongenotoxic carcinogens. p. 129–140, In: 1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice. P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Dutka, B.J. and G. Bitton, 1986. *Toxicity testing using microorganisms. Vol. 2.* CRC Press, Boca Raton, Fla.

Dutka, B.J., K. Jones, K.K. Kwan, H. Bailey, and R. McInnis, 1988. Use of microbial and toxicant screening tests for priority site selection of degraded areas in water bodies. *Water Res., 22:*503–510.

Eagleson, K.W., D.L. Lenat, L.W. Ausley, and F.B. Winborne, 1990. Comparison of measured instream biological responses with responses predicted using the *Ceriodaphnia dubia* chronic toxicity test. *Environ. Toxicol. Chem., 9:* 1019–1028.

EC (Environment Canada), 1971. Pulp and paper effluent regulations. Regulations, Codes and Protocols Rept 1, Environment Canada, Environmental Protection Service, Ottawa, Ont. 7 p.

———. 1972. Guidelines for the pulp and paper effluent regulations. Regulations, Codes and Protocols Rept 2, Environment Canada, Environmental Protection Service, Ottawa, Ont. 21 p.

———1990a. Biological test method: Acute lethality test using rainbow trout. EPS 1/RM/9, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1990b. Biological test method: Acute lethality test using threespine stickleback. EPS 1/RM/10, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1990c. Biological test method: Acute lethality test using *Daphnia* sp. EPS 1/RM/11, Environment Canada, Environmental Protection Service, Ottawa, Ont.

——— 1990d. Guidance document on control of toxicity test precision using reference toxicants, EPS 1/RM/12, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1990e. Reference method for determining acute lethality of effluents to rainbow trout. EPS 1/RM/13, Environment Canada. Environmental Protection Service, Ottawa, Ont.

———. 1990f. Reference method for determining acute lethality of effluents to *Daphnia magna*. EPS 1/RM/14, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1992a. Biological test method: Test of reproduction and survival using the cladoceran *Ceriodaphnia dubia*. EPS 1/RM/21, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1992b. Biological test method: Test of larval growth and survival using fathead minnows. Rept EPS 1/RM/22, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1992c. Biological test method: Toxicity test using luminescent bacteria *Photobacterium phosphoreum*. EPS 1/RM/24, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1992d. Biological test method: Growth inhibition test using the freshwater alga *Selenastrum capricornutum*. Rept EPS 1/RM/25, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1992e. Biological test method: Acute test for sediment toxicity using marine or estuarine amphipods. EPS 1/RM/26, Environment Canada, Environmental Protection service, Ottawa, Ont.

———. 1992f. Biological test method: Fertilization assay using echinoids (sea urchins and sand dollars). EPS 1/RM/27. Environment Cnaada, Environmental Protection service, Ottawa, Ont.

———. 1992g. Fertilization assay with echinoids: Interlaboratory evaluation of test options. Environment Canada, Conservation and Protection, Technology Development Branch, Ottawa, Ont. Unpub. Rept. 45 p. + app.

———. 1992h. Pulp and paper mill effluent chlorinated dioxins and furans regulations. Canada Gazette, Part II, Vol. 126, No. 11, May 20, 1992, p. 1940–1954.

———. 1993a. Canadian toxicological testing laboratory accreditation program. Program description. Environment Canada, Environ. Technol. Advancement Dir., Ottawa, Ont. 26 p.

———. 1993b. A discussion paper on ocean disposal. Environment Canada, Environ. Prot., Marine Environ. Div., Ottawa, Ont. 22 p.

———. 1993c. Interim monitoring guidelines for ocean disposal. Environment Canada, Environmental Protection, Marine Environ. Div., Ottawa, Ont. 34 p.

———. 1994a. Towards a toxic substances management policy for Canada. A discussion document for consultation purposes. Environment Canada, Ottawa, Ont. Sept. 1994, 23 + 29 p.

———. 1994b. Biological monitoring guidelines. Addendum to the interim monitoring guidelines for ocean disposal. Environment Canada, Environmental Protection, Marine Environ. Div., Ottawa, Ont. 6 p.

———. 1994c. Guidance document for the collection and preparation of sediment for physicochemical characterization and biological testing. EPS 1/RM/29, Environment Canada, Environ. Protection Service, Ottawa, Ont.

———. 1994d. A review of whole organism bioassays for assessing the quality of soil, freshwater sediment, and fresh water in Canada. Sc. Ser. Rept no. 198, Prepared by C. Keddy, J.C. Greene, and M.A. Bonnell for Environment Canada, Evaluation and Interpretation Branch, Ottawa, Ont. 185 p.

———. 1994e. A framework for ecological risk assessment at contaminated sites in Canada: Review and recommendations. Sc. Ser. no. 199, Prepared by C. Gaudet, EVS Environment Consultants, and ESSA Environmental and Social Systems Analysts for Environment Canada, Ecosystem Conservation Directorate, Evaluation and Interpretation Branch, Ottawa, Ont., 108 p.

———. 1994f. Interim guidance for the application of rapidly rendered harmless and trace contaminants. Unpublished document, November 1994, Environment Canada, Marine Environment Div., Ottawa, Ont. 6 p.

———. 1995a. Guidance document on measurement of toxicity test precision using control sediments spiked with a reference toxicant. EPS 1/RM/30, Environment Canada, Environmental Protection Service, Ottawa, Ont.

———. 1995b. Users guide to the application form for ocean disposal. EPS 1/MA/1, Environment Canada, Environmental Protection Service, Marine Environment Div., Ottawa, Ont.

———. 1995c. Interim sediment quality guidelines. Environment Canada, Guidelines Division, Evaluation and Interpretation Branch, Ottawa, Ont.

———. 1996a. Application of recommended whole organism bioassays in the assessment of contaminated sites in Canada. Prepared by C. Keddy, J.C. Greene, and M.A. Bonnell for Environment Canada, Evaluation and Interpretation Branch, Ottawa. Ont.

———. 1996b. 1996 national compendium monitoring at ocean disposal sites in Canada. Unnumbered rept, Environment Canada, Marine Environment Div., Ottawa, Ont., 36 p.

———. 1996c. Interim data interpretation guidelines for monitoring at ocean disposal sites. Unnumbered rept., Environment Canada, Marine Environment Div., Ottawa, Ont., 2 p.

———. 1997a. Biological test method. Test for growth and survival in sediment using larvae of freshwater midges *(Chironomus tentans* or *Chironomus riparius)*. EPS 1/RM/32, Environment Canada, Environment Protection Service, Ottawa, Ont.

———. 1997b. Biological test method. Test for growth and survival in sediment using the freshwater amphipod *Hyalella azteca*. EPS 1/RM/33, Environment Canada, Environment Protection Service, Ottawa, Ont.

———. 1997c. Toxicology Expert Working Group report. EEM/1997/5, Environment Canada, Ottawa, Ont.

———. 1997d. Environmental assessments of priority substances under the Canadian Environmental Protection Act. Guidance Manual Version 1.0. EC, EPS/2/CC/3E, Environment Canada, Commercial Chemicals Evaluation Branch, Chemicals Evaluation Division, Ottawa, Ont.

———. 1998. Biological test method: toxicity tests using early life stages of rainbow trout. Second edition. EPS 1/RM/28, Environment Canada, Environment Protection Service, Ottawa, Ont.

———. 1999. Biological test method: Test for measuring the inhibition of growth using the freshwater macrophyte *Lemna minor*. EPS 1/RM/37, Environment Canada, Environment Protection Service, Ottawa, Ont.

———. 2000a. Pulp and paper aquatic environmental effects monitoring requirements. EEM/2000/1, Annex 1. Environment Canada, Ottawa, Ont.

———. 2000b. Pulp and paper technical guidance for aquatic environmental effects monitoring. EEM/2000/2, Environment Canada, Ottawa, Ont.

———. in prep., a. Guidance document on statistical methods for estimating endpoints of toxicity tests. Rept EPS 1/RM/??, Environment Canada, Environ. Protection Service, Ottawa, Ont.

———. in prep., b. Biological test method. Test for growth and survival in sediment using marine or estuarine polychaete worms. EPS 1/RM/??, Environment Canada, Environment Protection Service, Ottawa, Ont.

EC and HWC (Environment Canada and Health and Welfare Canada), 1993. Guidelines for the notification and testing of new substances: Chemicals and polymers. Pursuant to the New Substances Notification Regulations of *Canadian Environmental Protection Act*. Environment Canada, Conservation and Protection, and Health and Welfare Canada, Health Protection Branch, Ottawa, Ont., 220 p.

EEC (European Economic Community), 1992a. Development and validation of methods for evaluating chronic toxicity to freshwater ecosystems. Final report (summary). EEC, Environ. Res. Programme, Assessment of Risk Associated with Chemicals (Ecotoxicology). Brussels.

———. 1992b. Part C: Methods for the determination of ecotoxicity. Official Journal of the European Communities L 383 A, *Vol. 35*:163–235.

Eisenhart, C., M.W. Hastay, and W.A. Wallis, 1947. *Techniques of statistical analysis*. McGraw-Hill, New York.

Elliot, J.E., R.W. Butler, R.J. Norstrom, and P.E. Whitehead, 1989. Environmental contaminants and reproductive success of great blue herons *Ardea herodias* in British Columbia, 1986–87. *Environ. Pollut., 59:*91–114.

Elliot, J.E. and R.J. Norstrom, 1998. Chlorinated hydrocarbon contaminants and productivity of bald eagle populations on the Pacific coast of Canada. *Environ. Toxicol. Chem., 17:*1142–1153.

Elson, P.F., 1966. Impact on salmon of spraying insecticide over forests. Advances in Water Pollut. Res., *Proc. 3rd Internat. Conf., Munich, 1966, 1:*55–74.

———. 1974. Impact of recent economic growth and industrial development on the ecology of Northwest Miramichi Atlantic salmon *(Salmo salar). J. Fish. Res. Board Can., 31:*521–544.

Elzerman, A. (ed.), 1992. Structure-activity relationships for environmental applications. *Environ. Toxicol. Chem. 11:*885–992.

Emans, H.J.B., E.J. v.d. Plassche, J.H. Canton, P.C. Okkerman, and P.M. Sparenburg, 1991. Validation of some extrapolation methods used for effect assessment. *Environ. Toxicol. Chem., 12:*2139–2154.

Emery, V.L. Jr., D.W. Moore, B.R. Gray, B.M. Duke, A.B. Gibson, R.B. Wright, and J.D. Farrar, 1997. Development of a chronic sublethal sediment bioassay using the estuarine amphipod *Leptocheirus plumulosus* (Shoemaker). *Environ. Toxicol. Chem., 16:*1912–1920.

Ernst, W., K. Doe, P. Jonah, J. Young, G. Julien, and P. Hennigar, 1991. The toxicity of chlorothalonil to aquatic fauna and the impact of its operational use on a pond ecosystem. *Arch. Environ. Contam. Toxicol., 21:*1–9.

Evans, R.D., J.R. Wisniewski, and J. Wisniewski (eds.), 1997. The interactions between sediments and water. *Proceedings of the 7th Internat. Sympos. Kluwer Acad. Pub.*, Norwell, Mass., 739 p.

Fahmy, F.K. and K. Schiefer, 1980. Application of the toxic contributions concept to multi-stream effluents. *Proc. Internat. sympos. on principles for interpretation of results of testing procedures in ecotoxicology,* Sept. 30–Oct. 2, 1980, Valbonne, France. [Commission of the European Communities, Rept EUR 7549, p. 453–462.]

Felsot, A.S., M.A. Bhatti, G.I. Mink, and G. Reisenauer, 1996. Biomonitoring with sentinel plants to assess exposure of nontarget crops to atmospheric deposition of herbicide residues. *Environ. Toxicol. Chem., 15:*452–459.

Ferrier, V., L. Gauthier, C. Zoll-Moreux, and J. L'Haridon, 1998. Genotoxicity tests in amphibians – A review. p. 507–519, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* Wells, P.G., K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Fisher, J.W., R.A. Dilego, M.E. Putnam, J.M. Livingston, and D.L. Geiger, 1983. Biological monitoring of bluegill activity. *Water Resour. Bull., 19:*211–215.

Fisk, A.T., R.J. Norstrom, C.D. Cymbalisty, and D.C.G. Muir, 1998. Dietary accumulation and depuration of hydrophobic organochlorines: Bioaccumulation parameters and their relationship with the octanol/water partition coefficient. *Environ. Toxicol. Chem., 17:* 951–961.

Fitzgerald, D.G., K.A. Warner, R.P. Lanno, and D.G. Dixon, 1996. Assessing the effects of modifying factors on pentachlorophenol toxicity to earthworms: Applications of body residues. *Environ. Toxicol. Chem., 15:*2299–2304.

Fletcher, J.S., J.E. Nellessen, and T.G. Pfleeger, 1994. Literature review and evaluation of the EPA food-chain (Kenaga) nomogram, an instrument for estimating pesticide residues on plants. *Environ. Toxicol. Chem., 13:*1383–1392

Fogels, A. and J.B. Sprague, 1977. Comparative short-term tolerance of zebrafish, flagfish and rainbow trout to five poisons including reference toxicants. *Water Res., 11:*811–817.

Forbes, T.L., V.E. Forbes, A. Giessing, R. Hansen, and L.K. Kure, 1998. Relative role of pore water versus ingested sediment in bioavailability of organic contaminants in marine sediments. *Environ. Toxicol. Chem., 17:*2453–2462.

Forsyth, D.J., C.F. Hinks, and N.D. Westcott, 1994. Feeding by clay-colored sparrows on grasshoppers and toxicity of carbofuran residues. *Environ. Toxicol. Chem., 13:*781–788.

Fossi, M.C., C. Leonzio, and S. Focardi, 1991. Modulation of mixed-function oxidase activity in black-headed gulls living in anthropic environments: biochemical acclimatization or adaptation? *Environ. Toxicol. Chem., 10:*1179–1188.

Foth, H.D. and L.M. Turk, 1972. *Fundamentals of soil science.* 5th ed. John Wiley & Sons, New York, N.Y. 454 p.

Fry, F.E.J., 1947. *Effects of the environment on animal activity.* Univ. Toronto Studies, Biol. Series no. 55, Pub. Ont. Fish. Res. Lab. 68: 62 p.

Fu, G., H.E. Allen, and Y. Cao, 1992. The importance of humic acids to proton and cadmium binding in sediments. *Environ. Toxicol. Chem., 11:*1363–1372.

Gagnon, C., M. Arnac, and J-R. Brindle, 1992. Sorption interactions between trace metals (Cd and Ni) and phenolic substances on suspended clay minerals. *Water Res., 26:*1067–1072.

Gagnon, M.M., J.J. Dodson, P.V. Hodson, G. Van Der Kraak, and J.H. Carey, 1994. Seasonal effects of bleached kraft mill effluent on reproductive parameters of white sucker *(Catostomus commersoni)* populations of the St. Maurice River, Quebec, Canada. *Can. J. Fish. Aquat. Sci., 51:*337–347.

Galassi, S., L. Guzzella, M. Mingazzini, L. Viganò, S. Capri, and S. Sora, 1992. Toxicological and chemical characterization of organic micropollutants in River Po waters (Italy). *Water Res., 26:*19–27.

Gardner, M. and A. Gunn, 1989. The effect of natural ligands on trace metal partitioning. *Chemosphere, 19:*1251–1259.

Gaudet, C.L., D.A. Milne, M.P. Wong, and T.G.E. Nason, 1995. A framework for ecological risk assessment at contaminated sites in Canada. Part II. Recommendations. *Human and Ecol. Risk Assessment, 1:*207–230.

Geckler, J.R., W.B. Horning, T.M. Neiheisel, Q.H, Pickering, E.L. Robinson, and C.E. Stephan, 1976. Validity of laboratory tests for predicting copper toxicity in streams. USEPA, Ecol. Res. Ser., EPA-600/3-76-116. Washington, D.C.

Giesy, J.P. and P.M. Allred, 1985. *Replicability of aquatic multispecies test systems.* p. 187–247, In: *Multispecies toxicity testing.* Cairns, J. Jr. (ed.), Pergamon Press, New York, N.Y.

Giesy, J.P. and R.A. Hoke, 1989. Freshwater sediment toxicity bioassessment: rationale for species selection and test design. *J. Great Lakes Res., 15:*539–569.

Giesy, J.P., C.J. Rosiu, and R.L. Graney, 1990. Benthic invertebrate bioassays with toxic sediment and pore water. *Environ. Toxicol. Chem., 9:*233–248.

Gillesby, B.E. and T.R. Zacharewski, 1998. Exoestrogens: Mechanisms of action and strategies for identification and assessment. *Environ. Toxicol. Chem., 17:*3–14.

Girling, A.E., R.K. Markarian, and D. Bennett, 1992. Aquatic toxicity testing of oil products -- Some recommendations. *Chemosphere 24:*1469–1472.

Golder [Golder Associates Ltd.], 1996. Guidance manual for the use of toxicity assessment to develop site specific remediation criteria for oil and gas facilities. Golder Assoc. Ltd., Calgary. Report prepared for Canad. Assoc. of Petroleum Producers, Calgary, Alta.

Graney, R.L., J.P. Giesy, and J.R. Clark, 1995. Field studies. p. 257–305, In: *Fundamentals of aquatic toxicology.* 2nd ed. G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

Graney, R.L., J.H. Kennedy, and J.H. Rodgers, 1994. Aquatic mesocosm studies in ecological risk assessment. Spec. Pub. Soc. Environ. Toxicol Chem., Lewis Pub., Boca Raton, Fla. 736 p.

Gray, L.E. Jr., J. Ostby, C. Wolf, C. Lambright, and W. Kelce, 1998. The value of mechanistic studies in laboratory animals for the prediction of reproductive effects in wildlife: endocrine effects on mammalian sexual differentiation. *Environ. Toxicol. Chem., 17:*109–118.

Green, D.W.J., K.A. Williams, D.R.L. Hughes, G.A.R. Shaik, and D. Pascoe, 1988. Toxicity of phenol to *Asellus aquaticus* (L.) -- Effects of temperature and episodic exposure. *Water Res., 22:*225–231.

Green, R.H., J.M. Boyd, and J.S. Macdonald, 1993. Relating sets of variables in environmental studies: The sediment quality triad as a paradigm. *Environmetrics, 4:*439–457.

Greene, J.C., C.L. Bartels, W.J. Warren-Hicks, B.R. Parkhurst, G.L. Linder, S.A. Peterson, and W.E. Miller, 1989. Protocols for short term toxicity screening of hazardous waste sites. USEPA, EPA-3/88/029.

Greim, H., J. Ahler, R. Bias, B. Broecker, A.O. Gamer, H.-P. Gelbke, W.G. Haltrich, H.-J. Klimisch, I. Mangelsdorf, N. Schön, G. Stropp, R. Vogel, G. Welter, and E. Bayer, 1993. Priority setting for the evaluation of existing chemicals; the approach of the German advisory committee on existing chemicals of environmental relevance (BUA). *Chemosphere, 26:*1653–1666.

Grothe, D. and R. Kimerle, 1985. Inter- and intra-laboratory variability in *Daphnia magna*. Effluent toxicity results. *Environ. Toxicol. Chem., 4:*189–192.

Grothe, D.R., K.L. Dickson, and D.K. Reed-Judkins (eds.), 1996. Whole effluent toxicity testing: an evaluation of methods and prediction of receiving system impacts. Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla. 346 p.

Guillette, L.J. Jr., D.B. Pickford, D.A. Crain, A.A. Rooney, and H.F. Percival, 1996. Reduction in penis size and plasma testosterone concentrations in juvenile alligators living in a contaminated environment. *Gen. Comp. Endocrinol., 101:*680–688.

Hall, A.T. and J.T. Oris, 1991. Anthracene reduces reproductive potential and is maternally transferred during long-term exposure in fathead minnows. *Aquatic Toxicol., 19:*249–264.

Hall, L.H., E.L. Maynard, and L.B. Kier, 1989. Structure-activity relationship studies on the toxicity of benzene derivatives: III. Predictions and extension to new substituents. *Environ. Toxicol. Chem., 8:*431–436.

Hamilton, J.D., and R. Sutcliffe, 1997. *Ecological assessment of polymers. Strategies for product stewardship and regulatory programs.* Van Nostrand Reinhold, Toronto. 345 p.

Hamr, P., R.S. Kirby, P. Gillis, and K.E. Day, 1994. Development of methodologies for whole-sediment toxicity tests with benthic invertebrates. (a) Optimum age for initiation and duration of the *Hyalella azteca* survival and growth test. (b) Formulation of artificial sediment(s) for use in reference toxicity tests with benthic invertebrates. Unpub. Rept. Environment Canada, Conservation and Protection, Technology Development Directorate, Ottawa, Ont. 28 p. + Tables and Figures.

Hansen, D.J., W.J. Berry, J.D. Mahony, W.S. Boothman, D.M. Di Toro, D.L.Robson, G.T. Ankley, D. Ma, Q. Yan, and C.E. Pesch, 1996. Predicting the toxicity of metal-contaminated field sediments using interstitial concentration of metals and acid-volatile sulfide normalizations. *Environ. Toxicol. Chem., 15:*2080–2094.

Hansen, S.R. and R.R. Garton, 1982. Ability of standard toxicity tests to predict the effects of the insecticide diflubenzuron on laboratory stream communities. *Can. J. Fish. Aquat. Sci., 39:*1273–1288.

Harborne, J.B., 1982. *Introduction to ecological biochemistry*. Academic Press, New York, N.Y.

Harkey, G.A., P.F. Landrum, and S.J. Klaine, 1994. Comparison of whole-sediment, elutriate and pore-water exposures for use in assessing sediment-associated organic contaminants in bioassays. *Environ. Toxicol. Chem., 13:*1315–1329.

Harshbarger, J.C. and J.B. Clark, 1990. Epizootiology of neoplasms in bony fish of North America. *Science Tot. Environ., 94:*1–32.

Hawkins, W.E., W.W. Walker, and R.M. Overstreet, 1995. Carcinogenicity tests using aquarium fish. p. 421–446, In: *Fundamentals of aquatic toxicology*. 2nd ed. G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

Hayward, R.S., N.G. Reichenbach, L.A. Dickson, and T.J. Wildoner Jr., 1988. Variability among bluegill ventilatory rates for effluent toxicity biomonitoring. *Water Res., 22:*1311–1315.

Heckman, C.W., 1983. The recovery of the biotic community in a lotic freshwater habitat after extensive destruction by chlorine. *Int. Revue ges. Hydrobiol., 68:*207–226.

Hellawell, J.M., 1978. *Biological surveillance of rivers. A biological monitoring handbook.* Water Research Centre, Elder Way, Stevenage, Herts., SG1 1TH, England. 332 p.

Henderson, C., 1957. Application factors to be applied to bioassays for the safe disposal of toxic wastes. p. 31–37, In: *Biological problems in water pollution. Trans. seminar on biological problems in water pollution,* C.M. Tarzwell (ed.), R.A. Taft Sanitary Engng Center, Cincinnati, Ohio, Apr. 23-27, 1956. U.S. Public Health Service, Dept Health, Education, and Welfare, Cincinnati, Ohio.

Henderson, C. and C.M. Tarzwell, 1957. Bio-assays for control of industrial effluents. *Sewage Indust. Wastes, 29:*1002–1017.

Henig, R.M., 1979. Animal experimentation: the battle lines soften. *Bioscience, 29:*145–148, 195–196.

Herbert, A, and P.-D. Hansen, 1998. Genotoxicity in fish embryos. p. 491–505, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* Wells, P.G., K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Hermens, J., H. Canton, P. Janssen, and R. De Jong, 1984a. Quantitative structure-activity relationships and toxicity studies of mixtures of chemicals with anaesthetic potency: acute lethal and sublethal toxicity to *Daphnia magna. Aquatic Toxicol., 5:*143–154.

Hermens, J., H. Canton, N. Steyger and R. Wegman, 1984b. Joint effects of a mixture of 14 chemicals on mortality and inhibition of reproduction of *Daphnia magna. Aquatic Toxicol., 5:*315–322.

Hersh, C.M. and W.G. Crumpton, 1989. Atrazine tolerance of algae isolated from two agricultural streams. *Environ. Toxicol. Chem., 8:*327–332.

Hicks, D.B. and J.W. DeWitt, 1971. Effects of dissolved oxygen on kraft pulp mill effluent toxicity. *Water Res., 5:* 693–701.

Ho, K.T.Y and J.G. Quinn, 1993. Physical and chemical parameters of sediment extraction and fractionation that influence toxicity, as evaluated by Microtox. *Environ. Toxicol. Chem., 12:*615–625.

Hobson, J.F. and W.J. Birge, 1989. Acclimation-induced changes in toxicity and induction of metallothionein-like proteins in the fathead minnow following sublethal exposure to zinc. *Environ. Toxicol. Chem., 8:*157–169.

Hodgson, E. and P.A. Levi, 1987. *A textbook of modern toxicology.* Elsevier Science Publishing Co., Inc. New York. 365 p.

Hodson, P.V., P.J. Kloepper-Sams, K.R. Munkittrick, W.L. Lockhart, D.A. Metner, P.L. Luxon, I.R. Smith, M.M. Gagnon, M. Servos, and J.F. Payne, 1991. Protocols for measuring mixed function oxygenases of fish liver. *Can. Tech. Rept Fish. Aquat. Sci., 1829,* 51 p.

Hodson, P.V., M.K. Maj, S. Efler, B.K. Burnison, A.R.P. van Heiningen, R. Girard, and J.H. Carey, 1997. MFO induction in fish by spent cooking liquors from kraft pulp mills. *Environ. Toxicol. Chem., 16:*908–916.

Hodson, P.V., J. Sherry, and J. Parrott, 1998. Bioassays to measure MFO inducers in effluents. p. 53–76, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* Wells, P.G., K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Hoekstra, J.A. and P.H. Van Ewijk, 1993. Alternatives for the no-observed-effect level. *Environ. Toxicol. Chem., 12:*187–194.

Hoffman, D.J., B.A. Rattner, G.A. Burton Jr, and J. Cairns Jr., 1995. *Handbook of ecotoxicology.* Lewis Publishers, Boca Raton, Fla. 768 p.

Holdway, D.A., E.J. Kempe, and D.G. Dixon, 1987. Acute toxicity of methoxychlor to larval white sucker *(Catostomus commersoni)* as modified by age, food availability, and egg preexposure. *Can. J. Fish. Aquat. Sci., 44:* 227–232.

Holloran, M.F., R.F. Willes, M.J. Palmiere, and D.G. Wilson, 1992. Estimating effluent toxicity from molecular-chlorine-gas-free bleaching in terms of toxicity equivalence. *Wat. Poll. Res. J. Can., 27:*539–559.

Hose, J.E., 1994. Large-scale genotoxicity assessments in the marine environment. *Environ. Health Perspect., 102 (Suppl. 12):*29–32.

Howard, D.E. and R. D. Evans, 1993. Acid-volatile sulphide (AVS) in a seasonally anoxic mesotrophic lake: seasonal and spatial changes in sediment AVS. *Environ. Toxicol. Chem., 12:*1051–1057.

Howard, T.E., 1978. Toxicity standards as modifiers of industrial operations. p. 58–62, In: *Proc. Fifth Annual Aquatic Toxicity Workshop,* Nov. 7-9, 1978, Hamilton, Ont. Canada, *Fish. Mar. Serv., Tech. Rept no. 862.*

Howarth, R.S. and J.B. Sprague, 1978. Copper lethality to rainbow trout in waters of various hardness and pH. *Water Res., 12:*455–462.

Hunt, D.T.E., [illegible]son, and R. Milne, 1992. The control and monitoring of discharges by biological techniques. *J. Inst. Water Environ. Management, 6:*269–277.

Hunt, J.W., B.S. Anderson, S.L. Turpen, M.A. Englund, and W. Piekarski, 1997. Precision and sensitivity of a 7-day growth and survival toxicity test using the west coast marine mysid crustacean *Holmesimysis costata. Environ. Toxicol. Chem., 16:*824–834.

Ikemoto, Y., K. Motoba, T. Suzuki, and M. Uchida, 1992. Quantitative structure-activity relationships of nonspecific and specific toxicants in several organism species. *Environ. Toxicol. Chem., 11:*931–940.

IMO (International Maritime Organization), 1990. *Intergovernmental Conference on the Convention on the Dumping of Wastes at Sea,* London, 30 October–13 November 1972. Final Act of the Conference with attachments including the Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matter. IMO, London. 1982 ed., reprinted 1990.

Ingersoll, C.G., 1995. Sediment tests. p. 231–255, In: *Fundamentals of aquatic toxicology.* 2nd ed. G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

Ingersoll, C.G., T. Dillon, and G.R. Biddinger (eds.), 1997. Ecological risk assessment of contaminated sediments. Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla. 389 p.

Ingersoll, C.G., E.L. Brunson, F.J. Dwyer, D.K. Hardesty, and N.E. Kemble, 1998. Use of sublethal endpoints in sediment toxicity tests with the amphipod *Hyalella azteca*. *Environ. Toxicol. Chem., 17:*1508–1523.

Inhaber, H., 1975. An approach to a water quality index for Canada. *Water Res., 9:*821–833.

Irwin, W.H., 1965. Fifty-seven species of fish in oil-refinery waste bioassay. *Trans. 30th North Amer. Wildl. Natural Resource Conf., 30:*89–99.

Isom, B.G., 1991. Biomonitoring. *Res. J. Water Pollut. Control Fed., 63:*746–748.

Jacobs, M.W., J.J. Delfino, and G. Bitton, 1992. The toxicity of sulphur to Microtox from acetonitrile extracts of contaminated sediments. *Environ. Toxicol. Chem., 11:*1137–1143.

Janssen, R.P.T., W.J.G.M. Peijnenburg, L. Posthuma, and M.A.G.T. Van Den Hoop, 1997a. Equilibrium partitioning of heavy metals in Dutch field soils. I. Relationship between metal partition coefficients and soil characteristics. *Environ. Toxicol. Chem., 16:*2470–2478.

Janssen, R.P.T., L. Posthuma, R. Baerselman, H.A. Den Hollander, R.P.M. Van Veen, and W.J.G.M. Peijnenburg, 1997b. Equilibrium partitioning of heavy metals in Dutch field soils. II. Prediction of metal accumulation in earthworms. *Environ. Toxicol. Chem., 16:*2479–2488.

Jaylet, A., P. Deparis, V. Ferrier, S. Grinfeld, and R. Siboulet, 1986. A new micronucleus test using peripheral blood erythrocytes of the newt *Pleurodeles waltl* to detect mutagens in fresh-water pollution. *Mutat. Res., 164:*245–257.

Jiminez, B.D. and L.S. Burtis, 1988. Response of the mixed-function oxidase system to toxicant dose, food and acclimation temperature in bluegill sunfish. *Mar. Environ. Res., 24:*45–49.

Jiminez, B.D., A. Oikari, S.M. Adams, D.E. Hinton, and J.F. McCarthy, 1990. Hepatic enzymes as biomarkers: Interpreting the effects of environmental, physiological and toxicological variables. p. 123–142, In: *Biomarkers of environmental contamination.* J.F. McCarthy, and L.R. Shugart (eds.), Lewis Pub., Boca Raton, Fla.

Johnson, B.T., 1998. Microtox® toxicity test system – new developments and applications. p. 201–218, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Jop, K.M., T.Z. Kendall, A.M. Askew, and R.B. Foster, 1991. Use of fractionation procedures and extensive chemical analysis for toxicity identification of a chemical plant effluent. *Environ. Toxicol. Chem., 10:*981–990.

Jung, K. and G. Bitton, 1997. Use of Ceriofast™ for monitoring the toxicity of industrial effluents: Comparison with the 48-h acute *Ceriodaphnia* toxicity test and Microtox®. *Environ. Toxicol. Chem., 16:*2264–2267.

Kasperson, R.E. and P.J.M. Stallen (eds.), 1991. *Communicating risks to the public. International perspectives.* Kluwer Academic Pub., Dordrecht, the Netherlands, 481 p.

Keith, L.H., 1997. *Environmental endocrine disruptors.* John Wiley & Sons, New York, N.Y.

Kendall, R., R. Dickerson, J. Giesy, and W. Suk (eds.), 1998. Principles and processes for evaluating endocrine disruption in wildlife. Soc. Environ. Toxicol. Chem., SETAC Press, Pensacola, Fla., 515 p.

Kersting, K., 1984. Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol., 69:*567–607.

Kimball, K.D. and S.A. Levin, 1985. Limitations of laboratory bioassays: The need for ecosystem-level testing. *Bioscience, 35:*165–171.

Klapow, L.A. and R.H. Lewis, 1979. Analysis of toxicity data for California marine water quality standards. *J. Water Pollut. Control Fed., 51:*2054–2070.

Kleinow, K.M., M.J. Melancon, and J.J. Lech, 1987. Biotransformation and induction: Implications for toxicity, bioaccumulation and monitoring of environmental xenobiotics in fish. *Environ. Health Perspect., 71:*105–119.

Klemm, D.J., L.B. Lobring, and W.H. Horning, 1990. Manual for the evaluation of laboratories performing aquatic toxicity tests. USEPA, Environmental Monitoring Systems Laboratory – Cincinnati, Cincinnati, Ohio, EPA/600/4-90-031. 109 p.

Knezovich, J.P., D.J. Steichen, J.A. Jelinski, and S.L. Anderson, 1996. Sulfide tolerance of four marine species used to evaluate sediment and porewater toxicity. *Bull Env. Contam. Toxicol., 57:*450–457.

Kocan, R.M., M.L. Landolt, and K.M. Sabo, 1982. Anaphase aberrations: A measure of genotoxicity in mutagen-treated fish cells. *Envir. Mutag., 4:*181–189.

Kooijman, S.A.L.M., 1985. Toxicity at population level. p. 143–165, In: *Multispecies toxicity testing*. Cairns, J. Jr. (ed.), Pergamon Press, New York, N.Y.

———. 1987. A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Res., 21:*269–276.

Koistinen, J., M. Soimasuo, K. Tukia, A. Oikari, A. Blankenship, and J.P. Giesy, 1998. Induction of EROD activity in Hepa-1 mouse hepatoma cells and estrogenicity in MCF-7 human breast cancer cells by extracts of pulp mill effluents, sludge, and sediment exposed to effluents. *Environ. Toxicol. Chem., 17:*1499–1507.

Kovacs, T. and B. O'Connor, 1996. Insights for toxicity-free pulp and paper mill effluents. Pulp and Paper Res. Inst. Canada, Pointe Claire (Quebec). Misc. Report MR 331, 26 p.

Kovacs, T.G., P.H. Martel, R.H. Voss, P.E. Wrist, and R.F. Willes, 1993. Aquatic toxicity equivalency factors for chlorinated phenolic compounds present in pulp mill effluents. *Environ. Toxicol. Chem., 12:*281–289.

Kovacs, T.G., R.H. Voss, S.R. Megraw, and P.H. Martel, 1996. Perspectives on the potential of pulp and paper mill effluents to affect the reproductive capacity of fish: A review of Canadian field studies. Pulp and Paper Res. Inst. of Canada, Pointe-Claire (Quebec), Misc. Rept MR 332, 41 p.

Krahn, M.M., M.S. Myers, D.G. Burrows, and D.C. Malins, 1984. Determination of metabolites of xenobiotics in the bile of fish from polluted waterways. *Xenobiotica, 14:*633–646.

Krahn, M.M., L.D. Rhodes, M.S. Myers, L.K. Moore, W.D. MacLeod Jr., and D.C. Malins, 1986. Associations between metabolites of aromatic hydrocarbons in bile and the occurrence of hepatic lesions in English sole *(Parophrys vetulus)* from Puget Sound, Washington. *Arch. Environ. Contam. Toxicol., 15:*61–67.

Krantzberg, G. and D. Boyd, 1992. The biological significance of contaminants in sediment from Hamilton harbour, Lake Ontario. *Environ. Toxicol. Chem., 11:*1527–1540.

Kszos, L.A., A.J. Stewart and P.A. Taylor, 1992. An evaluation of nickel toxicity to *Ceriodaphnia dubia* and *Daphnia magna* in a contaminated stream and in laboratory tests. *Environ. Toxicol. Chem., 11:*1001–1012.

Kszos, L.A., A.J. Stewart, and J.R. Sumner, 1997. Evidence that variability in ambient fathead minnow short-term chronic tests is due to pathogenic infection. *Environ. Toxicol. Chem., 16:*351–356.

Kuiper, J., 1981. Fate and effects of cadmium in marine plankton communities in experimental enclosures. *Marine Ecol. Progr. Ser., 6:*161–174.

Kukkonen, J, and P.F. Landrum, 1995. Measuring assimilation efficiencies for sediment-bound PAH and PCB congeners by benthic organisms. *Aquat. Toxicol. 32:*7—92. Errata, 1996. *Aquatic Toxicol., 34:*86.

———. 1996. Distribution of organic carbon and organic zenobiotics among different particle size fractions in sediments. *Chemosphere, 32:*1063–1076.

Kulakkattolickal, A.T. and D.L. Kramer, 1987. Effect of surface access and oxygen concentration on the toxicity of *Croton tiglium* (Euphorbiaceae) seed extract to *Brachydanio rerio* (Cyprinidae). *Can. J. Fish. Aquat. Sci., 44:*1358–1361.

Kumaraguru, A.K. and F.W.H. Beamish, 1981. Lethal toxicity of Permethrin (NRDC-143) to rainbow trout, *Salmo gairdneri*, in relation to body weight and water temperature. *Water Res., 15:*503–505.

Lamberson, J.O., T.H. DeWitt, and R.C. Swartz, 1992. Assessment of sediment toxicity to marine benthos. p. 183–287, In: 1992. *Sediment toxicity assessment*. G.A. Burton, Jr. (ed.), Lewis Pub., Boca Raton, Fla.

Landis, W.G., and M.-H. Yu, 1998. *Introduction to environmental toxicology*. 2nd ed. Lewis Pub., Boca Raton, Fla. 400 p.

Landrum, P.F. and W.R. Faust, 1995. The role of sediment composition on the bioavailability of laboratory-dosed sediment-associated organic contaminants to the amphipod *Diporeia spp. Chem. Speciat. Bioavail. 6:85–92.*

Langevin, R., J.B. Rasmussen, H. Sloterdijk, and C. Blaise, 1992. Genotoxicity in water and sediment extracts from the St Lawrence River system, using the SOS Chromotest. *Water Res., 26:*419–429.

La Point, T.W., and J.F. Fairchild, 1992. Evaluation of sediment contaminant toxicity: the use of freshwater community structure. p. 87–110, In: *Sediment toxicity assessment*. G.A. Burton (ed.), Lewis Publishers. Boca Raton, Fla.

Larkin, P.A., 1984. A commentary on environmental impact assessment for large projects affecting lakes and streams. *Can. J. Fish. Aquat. Sci., 41:*1121–1127.

Larsen, D.P., F. deNoyelles, F. Stay, and T. Shiroyama, 1986. Comparisons of single species, microcosm and experimental pond responses to atrazine exposure. *Environ. Toxicol. Chem., 5:*179–190.

Lazorchak, J.M. and W.T. Waller, 1993. The relationship of total copper 48-h LC50s to *Daphnia magna* dry weight. *Environ. Toxicol. Chem., 12:*903–912.

Leach, J.M. and A.N. Thakore, 1977. Compounds toxic to fish in pulp mill waste streams. *Prog. Water Tech., 9:* 787–798.

LeBlanc, G.A., 1984. Interspecies relationships in acute toxicity of chemicals to aquatic organisms. *Environ. Toxicol. Chem., 3:*47–60.

Lee, K. and K.L. Tay, 1998. Measurement of microbial exoenzyme activity in sediments for environmental impact assessment. p. 219–236, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Legault, G. and M. Villeneuve, 1992. Chimiotox: results of a chemo-toxic evaluation of the industrial plants targeted under the St. Lawrence Action Plan. Volume 1. L'Equipe d'intervention St. Laurent, Montréal, Québec. 20 p.

Legault, R., C. Blaise, D. Rokosh, and R. Chong-Kit, 1994. Comparative assessment of the SOS Chromotest kit and the Mutatox test with the *Salmonella* Plate Incorporation (Ames Test) and fluctuation tests for screening genotoxic agents. *Environ. Toxicol. and Water Qual., 9:*45–47.

Lehman, R.M., F.S. Colwell, and J.L. Garland, 1997. Physiological profiling of indigenous aquatic microbial communities to determine toxic effects of metals. *Environ. Toxicol. Chem., 16:*2232–2241.

Leiss, W. (ed.), 1989. *Prospects and problems in risk communication.* Univ. Waterloo Press, Waterloo, Ont., 216 p.

Leopold, L.B., F.E. Clarke, B.B. Hanshaw, and J.R. Balsley, 1971. A procedure for evaluating environmental impact. U.S. Geological Survey, Washington, D.C., Circular 645.

Leppanen, C.J. and K.J. Maier, 1998. An inexpensive and efficient modular water-renewal system for bulk sediment toxicity testing. *Environ. Toxicol. Chem., 17:*969–971.

Levin, S.A. and K.D. Kimball (eds.), 1984. New perspectives in ecotoxicology. (Various papers included) *Environ. Management, 8:*375–442.

Lewis, M.A., R.L. Powell, M.K. Nelson, M.G. Henry, S.J. Klaine, G.W. Dickson, and F.L. Mayer, 1999. Ecotoxicology and risk assessment for wetlands. Soc. Environ. Toxicol. Chem., SETAC Press, Pensacola, Fla.

L'Haridon, J., M. Fernandez, V. Ferrier, and J. Bellan, 1993. Evaluation of the genotoxicity of *N*-Nitrosoatrazine, *N*-Nitrosodiethanolamine and their precursors *in vivo* using the newt micronucleus test. *Water Res., 27:*855–862.

Liber, K., D.J. Call, T.P. Markee, K.L. Schmude, M.D. Balcer, F.W. Whiteman, and G.T. Ankley, 1996. Effects of acid-volatile sulfide on zinc bioavailability and toxicity to benthic macroinvertebrates: a spiked-sediment field experiment. *Environ. Toxicol. Chem., 15:*2113–2125.

Lipnick, R.L., 1995. Structure-activity relationships. p. 609–655, In: *Fundamentals of aquatic toxicology.* 2nd ed. G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

Lloyd, R., 1961. Effect of dissolved oxygen concentrations on the toxicity of several poisons to rainbow trout *(Salmo gairdnerii* Richardson). *J. Exp. Biol., 38:*447–455.

Løkke, H., and C.A.M. van Gestel (eds.), 1998. *Handbook of soil invertebrate toxicity tests.* John Wiley & Sons Ltd., New York, New York, 281 p.

Long, E.R., 1992. Ranges in chemical concentrations in sediments associated with adverse biological effects. *Mar. Poll. Bull., 24:*38–45.

Long, E.R. and D.D. MacDonald, 1992. National status and trends program approach. p. 1–18, In: *Sediment classification methods compendium.* USEPA, Washington, D.C, EPA 823-R-92-006.

Long, E.R. and L.G. Morgan, 1990. The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program. Nat. Oceanic and Atmospheric Admin., Tech. Memorandum NOS OMA 52. Seattle, Wash., 175 p. + app.

Long, E.R., M.F. Buchman, S.M. Bay, R.J. Breteler, R.S. Carr, P.M. Chapman, J.E. Hose, A.L. Lissner, J. Scott, and D.A. Wolfe, 1990. Comparative evaluation of five toxicity tests with sediments from San Francisco Bay and Tomales Bay, California. *Environ. Toxicol. Chem,. 9:*1193–1214.

Luoma, S.N. and J.L. Carter, 1993. Understanding the toxicity of contaminants in sediments: beyond the bioassay-based paradigm. *Environ. Toxicol. Chem., 12:*793–796.

MacDonald, D.D., L.E. Fidler, and D. Valiela, 1987. Site specific water quality criteria for fish and aquatic life in the Canadian portion of the Flathead River basin. Nitrate, nitrite and ammonia. Unnum. rept, Environment Canada, Conservation and Protection, Inland Waters and Lands, Pacific and Yukon Region. Vancouver, B.C. 127 p.

MacDonald, D.D., S.L. Smith, M.P. Wong, and P. Murdoch, 1992. The development of Canadian marine environmental quality guidelines. Prepared for the Interdepartmental Working Group on Marine Environmental Quality Guidelines and the Canadian Council of Ministers of the Environment. Environment Canada, Ottawa, Ont., 50 p. + app.

Mahony, J.D., D.M. Di Toro, A.M. Gonzalez, M. Curto, M. Dilg, L.D. De Rosa, and L.A. Sparrow, 1996. Partitioning of metals to sediment organic carbon. *Environ. Toxicol. Chem., 15:*2187–2197.

Malins, D.C., B.B. McCain, D.W. Brown, S. Chan, M.S. Myers, J.T. Landahl, P.G. Prohaska, A.J. Friedman, L.D. Rhodes, D.G. Burrows, W.D. Gronlund, and H.O. Hodgins, 1984. Chemical pollutants in sediments and diseases of bottom-dwelling fish in Puget Sound, Washington. *Envir. Sci. Tech., 18:*705–713.

Marcus, M.D. and L.L. McDonald, 1992. Evaluating the statistical bases for relating receiving water impacts to effluent and ambient toxicities. *Environ. Toxicol. Chem., 11:*1389–1402.

Marmonier, P., P. Vervier, J. Gibert, and M-J. Dole-Olivier, 1993. Biodiversity in groundwaters. *TREE, 8:*392–395.

Maron, D.M. and Ames, B.N, 1983. Revised methods for the *Salmonella* mutagenicity test. *Mutation Res., 113:* 173–215.

Martel, P. and T. Kovacs, 1997. A comparison of the potential of primary- and secondary-treated pulp and paper mill effluents to induce mixed function oxygenase (MFO) activity in fish. *Wat. Res., 31:*1482–1488.

Martel, P.H., T.G. Kovacs, B.I. O'Connor, and R.H. Voss, 1997. Source and identity of compounds in a thermomechanical pulp mill effluent inducing hepatic mixed-function oxygenase activity in fish. *Environ. Toxicol. Chem., 16:*2375–2383.

Marty, G.D., S. Wetzlich, J.M. Nunez, A. Craigmill, and D.E. Hinton, 1991. Fish-based biomonitoring to determine toxic characteristics of complex chemical mixtures: documentation of bioremediation at a pesticide disposal site. *Aquat. Tox., 19:*329–340.

Matheson, Robert A. and A.J. Atkinson, 1997. Regulation of new polymers in Canada. p. 235–253, In: *Ecological assessment of polymers. Strategies for product stewardship and regulatory programs.* J.D. Hamilton and R. Sutcliffe (eds.), Van Nostrand Reinhold, Toronto, Ont.

Matthiessen, P. and P.E. Gibbs, 1998. Critical appraisal of the evidence for tributyltin-mediated endocrine disruption in mollusks. *Environ. Toxicol. Chem., 17:*37–43.

Mayer, F.L., P.M. Mehrle, and P.L. Crutcher, 1978. Interactions of toxaphene and vitamin C in channel catfish. *Trans. Amer. Fish. Soc., 107:*326–333.

McCarty, L.S., 1991. Toxicant body residues: implications for aquatic bioassays with some organic chemicals. p. 183–192, In: *Aquatic toxicology and risk assessment.* 14th vol. M.A. Mayes and M.G. Barron (eds.), Amer. Soc. for Testing and Materials, Philadelphia, Pa., ASTM STP 1124.

McCarty, L.S. and D. Mackay, 1993. Enhanced ecotoxicological modeling and assessment: Body residues and modes of toxic action. *Environ. Science Technol., 27:*171–1727.

McCarty, L.S., D. Mackay, A.D. Smith, G.W. Ozburn, and D.G. Dixon, 1992. Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: neutral narcotic agents. *Environ. Toxicol. Chem., 11:* 917–930.

McCloskey, J.T. and J.T. Oris, 1991. Effect of water temperature and dissolved oxygen concentration on the photo-induced toxicity of anthracene to juvenile bluegill sunfish *(Lepomis macrochirus). Aquatic Toxicol., 21:*145–156.

McCubbin, N. and J. Folke, 1993. Environmental Choice Program guidelines for pulp and paper products. Briefing notes for pulp and paper products. Prepared for Environment Canada and Forestry Canada, Environmental Choice Program, Ottawa, Ont., 140 p.

McGeachy, S.M. and D.G. Dixon, 1990. Effect of temperature on the chronic toxicity of arsenate to rainbow trout *(Oncorhynchus mykiss). Can. J. Fish. Aquat. Sci., 47:*2228–2234.

McKim, J.M., S.P. Bradbury, and G.J. Niemi, 1987. Fish acute toxicity syndromes and their use in the QSAR approach to hazard assessment. *Environ. Health Perspect., 71:*171–186.

McLeay, D.J., C.C. Walden, and J.R. Munro, 1979. Effect of pH on toxicity of kraft pulp and paper mill effluent to salmonid fish in fresh and seawater. *Water Res., 13:*249–254.

McMaster, M.E., G.J. Van Der Kraak, C.B. Portt, K.R. Munkittrick, P.K. Sibley, I.R. Smith, and D.G. Dixon, 1991. Changes in hepatic mixed-function oxygenase (MFO) activity, plasma steroid levels and age at maturity of a white sucker *(Catostomus commersoni)* population exposed to bleached kraft pulp mill effluent. *Aquatic Toxicol., 21:*199–218.

McMaster, M.E., C.B. Portt, K.R. Munkittrick, and D.G. Dixon, 1992. Milt characteristics, reproductive performance, and larval survival and development of white sucker exposed to bleached kraft mill effluent. *Ecotoxicol. Environ. Safety, 23:*103–117.

McMaster, M.E., G.J. Van Der Kraak, and K.R. Munkittrick, 1996. Exposure to bleached kraft pulp mill effluent reduces the steroid biosynthetic capacity of white sucker ovarian follicles. *Comp. Biochem. Physiol., C 112:*169–178.

Mehrle, P.M., F.L. Mayer, and W.W. Johnson, 1977. Diet quality in fish toxicology: effects on acute and chronic toxicity. p. 269–280, In: F.L. Mayer and J.L. Hamelink (eds.), Aquatic toxicology and hazard evaluation. ASTM STP 634. Amer. Soc. for Testing and Materials, Philadelphia, Pa.

Meier, J.R., L.W. Chang, S. Jacobs, J. Torsella, M.C. Meckes, and M.K. Smith, 1997. Use of plant and earthworm bioassays to evaluate remediation of soil from a site contaminated with polychlorinated biphenyls. *Environ. Toxicol. Chem., 16:*928–938.

MENVIC [Ministère de l'environnement], 1990. Cahier des critères de qualité de l'eau. Québec, Ministère de l'environnement, Québec (Québec).

Merritt, R.W. and K.W. Cummins, 1984. *An introduction to the aquatic insects of North America.* 2nd ed. Kendall-Hunt Pub. Co., Dubuque, Iowa.

Meyers-Schöne, L., L.R. Shugart, J.J. Beauchamp, and B.T. Walton, 1993. Comparison of two freshwater turtle species as monitors of radionuclide and chemical contamination: DNA damage and residue analysis. *Environ. Toxicol. Chem., 12:*1487–1496.

Microbics, 1992. *The Microtox on-line toxicity test system.* Microbics Corporation, Carlsbad, Calif., 4 p.

———. 1993. Mutatox™ manual. Microbics Corp., Carlsbad, Calif., 22 p.

Milani, D., K.E. Day, D.J. McLeay, and R.S. Kirby, 1996. Recent intra- and interlaboratory studies related to the development and standardization of Environment Canada's biological test methods for measuring sediment toxicity using freshwater amphipods *(Hyalella azteca)* or midge larvae *(Chironomus riparius)*. National Water Res. Inst. Technical Rept, Environment Canada, Method Development and Application Section, Ottawa, Ont., unnumbered.

Mix, M.D., 1986. Cancerous diseases in aquatic animals and their association with environmental pollutants: a critical review of the literature. *Mar. Fish. Rev., 20:*1–141.

Moore, D.R.J. and P.-Y. Caux, 1997. Estimating low toxic effects. Environ. *Toxicol. Chem., 16:*794–801.

Moore, D.W., T.S. Bridges, B.R. Gray, and B.M. Duke, 1997. Risk of ammonia toxicity during sediment bioassays with the estuarine amphipod *Leptocheirus plumulosus. Environ. Toxicol. Chem., 16:*1020–1027.

Moore, M.V. and R.W. Winner, 1989. Relative sensitivity of *Ceriodaphnia dubia* laboratory tests and pond communities of zooplankton and benthos to chronic copper stress. *Aquatic Toxicol., 15:*311–330.

Morrison, G.E., E. Torrelo, R. Comeleo, R. Walsh, A. Kuhn, R. Burgess, M. Tagliabue, and W. Greene, 1989. Intralaboratory precision of saltwater short-term chronic toxicity tests. *J. Water Pollut. Control. Fed., 61:*1707–1710.

Moss, D., M.T. Furse, J.F. Wright, and P.D. Armitage, 1987. The prediction of the macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwat. Biol., 17:*41–52.

Mothes-Wagner, U., H.K. Reitze, and K-A. Seitz, 1992. Terrestrial multispecies toxicity testing. 1. Description of the multispecies assemblage. *Chemosphere, 24:*1653–1667.

Mount, D.I. and T.J. Norberg-King (eds.), 1985. Validity of effluent and ambient toxicity tests for predicting biological impact, Scippo Creek, Circleville, Ohio. USEPA, Environ. Res. Lab., Duluth, Minn., and Permits Div., Washington, D.C. EPA/600/3-85/044.

———. 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Kanawha River, Charleston, West Virginia. USEPA, Environ. Res. Lab., Duluth, Minn., and Permits Div., Washington, D.C. EPA/600/3-86/006.

Mount, D.I., N.A. Thomas, T.J. Norberg, M.T. Barbour, T.H. Roush, and W.F. Brandes, 1984. Effluent and ambient toxicity testing and instream community response on the Ottawa River, Lima, Ohio. USEPA, Permits Div., Washington, D.C., and Environ. Res. Lab., Duluth, Minn. EPA-600/3-84-080.

Mount, D.I., A.E. Steen, and T.J. Norberg-King (eds.), 1985. Validity of effluent and ambient toxicity testing for predicting biological impact on Five Mile Creek, Birmingham, Alabama. USEPA, Environ. Res. Lab., Duluth, Minn., and Permits Div., Washington, D.C. EPA/600/8-85/015.

———. 1986a. Validity of effluent and ambient toxicity tests for predicting biological impact, Naugatuck River, Waterbury, Connecticut. USEPA, Environ. Res. Lab., Duluth, Minn., and Permits Div., Washington, D.C. EPA/600/8-86/005.

———. 1986b. Validity of effluent and ambient toxicity tests for predicting biological impact, Back River, Baltimore Harbor, Maryland. USEPA, Environ. Res. Lab., Duluth, Minn and Permits Div., Washington D.C. Rept EPA/600/8-86/001.

———. 1986c. Validity of ambient toxicity tests for predicting biological impact, Ohio River, near Wheeling, West Virginia. USEPA, Environ. Res. Lab., Duluth, Minn., and Permits Div., Washington, D.C. EPA/600/3-85/071.

Mudroch, A. and J.M. Azcue, 1995. *Manual of aquatic sediment sampling*. Lewis Pub., Boca Raton, Fla. 240 p.

Müller, W.E.G. and I.M. Müller, 1998. Sponge cells and tissue as biological monitors of aquatic pollution. p. 97–112, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice*. P.G. Wells, K. Lee, and C. Blaise (eds.) CRC Press, Boca Raton, Fla.

Munkittrick, K.R. and L.S. McCarty, 1995. An integrated approach to aquatic ecosystem health: Top-down, bottom-up or middle-out? *J. Aquat.Ecosystem Health, 4*:77–90.

Munkittrick, K.R., C.B. Portt, G.J. Van Der Kraak, I.R. Smith, and D.A. Rokosh, 1991. Impact of bleached kraft mill effluent on population characteristics, liver MFO activity, and serum steroid levels of a Lake Superior white sucker *(Catostomus commersoni)* population. *Can. J. Fish. Aquat. Sci., 48*:1371–1380.

Munkittrick, K.R., G.J. Van Der Kraak, M.E. McMaster, and C.B. Portt, 1992. Response of hepatic MFO activity and plasma sex steroids to secondary treatment of bleached kraft pulp mill effluent and mill shutdown. *Environ. Toxicol. Chem., 11*:1427–1439.

Nabholz, J.V., R.G. Clements, M.G. Zeeman, K.C. Osborn, and R. Wedge, 1993. Validation of structure activity relationships used by the USEPA's Office of Pollution Prevention and Toxics for the environmental hazard assessment of industrial chemicals. p. 571–590, In: Environmental toxicology and risk assessment. 2nd vol. J.W. Gorsuch, F.J. Dwyer, C.G. Ingersoll, and T.W. LaPoint (eds.), Amer. Soc. Testing and Materials, Philadelphia, Pa., ASTM STP 1216.

Nacci, D., L. Coiro, A. Kuhn, D. Champlin, W. Munns Jr., J. Specker, and K. Cooper, 1998. Nondestructive indicator of ethoxyresorufin-*o*-deethylase activity in embryonic fish. *Environ. Toxicol. Chem., 17*:2481–2486.

Naish, V.A., 1975. Red Rock toxicity study. Domtar Research Centre, Senneville, Quebec, Res. Rept, Proj. 73-6107-12, Rept No. 2, 16 p.

NATO (North Atlantic Treaty Organization), 1988. Pilot study on international information exchange on dioxins and related compounds. North Atlantic Treaty Organization, Committee on the Challenges of Modern Society, Brussells, Rept no. 177.

Naylor, C., L. Pindar, and P. Callow, 1990. Inter- and intraspecific variation in sensitivity to toxins; the effects of acidity and zinc on the freshwater crustaceans *Asellus aquaticus* (L.) and *Gammarus pulex* (L.). *Water Res., 24*: 757–762.

Naz, R.K. (ed.), 1998. *Endocrine disruptors. Effects on male and female reproductive systems*. CRC Press, Boca Raton, Fla., 352 p.

Nebeker, A.V., 1982. Evaluation of a *Daphnia magna* renewal life-cycle test method with silver and endosulfan. *Water Res., 16*:739–744.

Nebeker, A.V., G.S. Schuytema, W.L. Griffis, J.A. Barbitta, and L.A. Carey, 1989. Effect of sediment organic carbon on survival of *Hyalella azteca* exposed to DDT and endrin. *Environ. Toxicol. Chem., 8*:705–718.

Nelson, M.K., P.F. Landrum, G.A. Burton Jr., S.J. Klaine, E.A. Crecelius, T.D. Byl, D.C. Gossiaux, V.N. Tsymbal, L. Cleveland, C.G. Ingersoll, and G. Sasson-Brickson, 1993. Toxicity of contaminated sediments in dilution series with control sediments. *Chemosphere, 27:*1789–1812.

Newman, M.C., 1995. *Quantitative methods in aquatic ecotoxicology.* Lewis Pub., Boca Raton, Fla.

Nimmo, D.W.R., D. Link, L.P. Parrish, G.J. Rodriguez, W. Wuerthele, and P.H. Davies, 1989. Comparison of on-site and laboratory toxicity tests: derivation of site-specific criteria for un-ionized ammonia in a Colorado transitional stream. *Environ. Toxicol. Chem., 8:*1177–1189.

Noppert, F., N. van der Hoeven, and A. Leopold (eds.), 1994. How to measure no effect. Towards a new measure of chronic toxicity in ecotoxicology. Workshop Rept, The Hague, The Netherlands, Sept. 9, 1994. The Netherlands Working Group on Statistics and Ecotoxicology. [Copies: BKH Consulting Engineers, P.O. box 5094, 2600 GB, Delft, The Netherlands, att. F. Noppert.]

Norberg-King, T.J. and D.I. Mount (eds.), 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Skeleton Creek, Enid, Oklahoma. USEPA, Environ. Res. Lab., Duluth, Minn. EPA/600/8-86/002.

Norberg-King, T.J., E.J. Durhan, and G.T. Ankley, 1991. Application of toxicity identification procedures to the ambient waters of the Colusa Basin Drain, California. *Environ. Toxicol. Chem., 10:*891–900.

Notenboom, J., S. Plénet, and M.-J. Turquin, 1994. Groundwater contamination and its impact on groundwater animals. p. 477–504, In: J. Gibert, D.L. Danielopol, and J.A. Stanford (eds.), Groundwater ecology. Academic Press, Toronto, Ont.

NRA (National Rivers Authority), 1991. Proposals for statutory water quality objectives. Report of the National Rivers Authority. National Rivers Authority, Rivers House, Waterside Drive, Aztec West, Almondsbury, Bristol BS12 4UD, United Kingdom. NRA Water Quality Series No. 5, 99 p.

NTAC (National Technical Advisory Committee), 1968. Water quality criteria. Report of the National Technical Advisory Committee to the Secretary of the Interior. U.S. Federal Water Pollution Control Admin., Washington, D.C. 234 p.

Obst, U., A. Wessler, and M. Wiegand-Rosinus, 1998. p. 77–94, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

O'Connor, T.P., K.D. Daskalakis, J.L. Hyland, J.F. Paul, and J.K. Summers, 1998. Comparisons of sediment toxicity with predictions based on chemical guidelines. *Environ. Toxicol. Chem., 17:*468–471.

OECD (Organization for Economic Cooperation and Development), 1987. Utilisation des tests biologiques pour l'evaluation et le contrôle de la pollution de l'eau. OECD, Paris, Monographies sur l'Environnement, No. 11, 76 p.

———. 1992a. Report of the OECD workshop on quantitative structure-activity relationships (QSARS) in aquatic effects assessment. OECD, Paris, OECD Environment Monographs No. 58, OCDE/GD(92)168, 95 p.

———. 1992b. Report of the OECD workshop on the extrapolation of laboratory aquatic toxicity data to the real environment. OECD, Paris, OECD Environment Monographs No. 59. OCDE/GD(92)169, 43 p.

———. 1992c. Report of the OECD workshop on effects assessment of chemicals in sediment. OECD, Paris, OECD Environment Monographs No. 60., OCDE/GD(92)170, 56 p.

———. 1992d. OECD guidelines for testing of chemicals. Section 2: effects on biotic systems. Summary of considerations in the report from the OECD expert group on ecotoxicology. Subsections: Methods 201–210. OECD, Paris.

———1992e. The OECD principles of good laboratory practice. Environment Monograph no. 45. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 1. OECD, Paris, OCDE/GD(92)32.

———. 1992f. Guidance for GLP monitoring authorities. Guides for compliance monitoring procedures for good laboratory practice. Environment Monograph no. 46. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 2. OECD, Paris, OCDE/GD(92)33.

———. 1992g. Guidance for GLP monitoring authorities. Guidance for the conduct of laboratory inspections and study audits. Environment Monograph no. 47. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 3. OECD, Paris, OCDE/GD(92)34.

———. 1992h. GLP consensus document. Quality assurance and GLP. Environment Monograph no. 48. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 4. OECD, Paris, OCDE/GD(92)35.

———. 1992i. GLP consensus document. Compliance of laboratory suppliers with GLP principles. Environment Monograph no. 49. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 5. OECD, Paris, OCDE/GD(92)36.

———. 1992j. OECD guideline for testing of chemicals. Fish, acute toxicity test. OECD, Paris. Adopted by the Council, 17th July 1992, 13 p.

———. 1993a. Technical guidance documents in support of the risk assessment directive. (93/67/EEC) for substances notified in accordance with the requirements of Council Directive 67/548/EEC. Chapter 11. OECD, Paris. 51 p.

———. 1993b. GLP consensus document. The application of the GLP principles to short-term studies. Environment Monograph no. 73. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 7. OECD, Paris, OCDE/GD(93)104.

———. 1993c. GLP consensus document. The role and responsibilities of the study director in GLP studies. Environment Monograph no. 74. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 8. OECD, Paris, OCDE/GD(93)105.

———. 1993d. OECD guidelines for the testing of chemicals. Volume 2. OECD, Paris.

———. 1994. GLP consensus document. The application of the GLP principles to field studies. Environment Monograph no. 50. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 6. OECD, Paris, OCDE/GD(92)37.

———. 1995a. Guidance document for aquatic effects assessment. OECD, Paris. Environment Monographs No. 92, 116 p.

———. 1995b. Report of the OECD workshop on environmental hazard/risk assessment. OECD, Paris. Environment Monographs No. 105, 96 p.

———. 1998a. Report of the OECD workshop on statistical analysis of aquatic toxicity data. OECD, Paris. OECD Environmental Health and Safety Publications, Series on Testing and Assessment No. 10, 133 p.

———. 1998b. Detailed review paper on aquatic testing methods for pesticides and industrial chemicals. Part 1: report. Part 2: annexes. OECD, Paris. OECD Environmental Health and Safety Publications, Series on Testing and Assessment No. 11, 260 + 303 p.

Okkerman, P.C., E.J. Plassche, C.J. Roghair, and J.H. Canton, 1990. Validation of some extrapolation methods with toxicity data derived from multiple species experiments. Presented at *OECD Workshop on Ecotoxicological Effects Assessment*, Arlington, Va. Rijksinstituut voor Volksgezondheid en Milieuhygiene [National Inst. of Public Health and Environmental Protection], Bilthoven, the Netherlands, 14 p. + app.

OMOE (Ontario Ministry of the Environment), 1984. Water management. Goals, policies, objectives and implementation procedures of the Ministry of the Environment. Ont. Min. Environ., Toronto, Ont., Nov. 1978, revised May 1984. 70 p.

———. 1991. Ontario's water quality objective development process. Ont. Min. Environ., Aquatic Criteria Development Committee, Toronto, Ont., 67 p.

Ontario, 1975. The environmental assessment act, 1975. Statutes of Ontario, 1975, Chapt. 69.

Oris, J.T. and J.P. Giesy Jr., 1985. The photoenhanced toxicity of anthracene to juvenile sunfish *(Lepomis* spp.). *Aquatic Toxicol., 6:*133–146.

———. 1987. The photo-induced toxicity of polycyclic aromatic hydrocarbons to larvae of the fathead minnow. *Chemosphere, 16:*1395–1404.

Oris, J.T., A.T. Hall, and J.D. Tylka, 1990. Humic acids reduce the photo-induced toxicity of anthracene to fish and *Daphnia*. *Environ. Toxicol. Chem., 9:*575–583.

Orr, P.L., R.W. Bradley, J.B. Sprague, and N.J. Hutchinson, 1986. Acclimation-induced change in toxicity of aluminum to rainbow trout *(Salmo gairdneri)*. *Can. J. Fish. Aquat. Sci., 43:*243–246.

Ortego, L.S. and W.H. Benson, 1992. Effects of dissolved humic material on the toxicity of selected pyrethroid insecticides. *Environ. Toxicol. Chem., 11:*261–266.

Ostrander, G., 1997. *Techniques in aquatic toxicology.* Lewis Pub., Boca Raton, Fla. 704 p.

Overstreet, R.M., 1987. Aquatic pollution problems, southwestern U.S. coasts: Histopathological indicators. *Aquat. Toxicol., 11:*213–239.

OWRC (Ontario Water Resources Commission), 1970. Guidelines and criteria for water quality management in Ontario. OWRC, Toronto, Ont. 26 p.

Pack, S., 1993. A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines. Shell Research Ltd., Sittingbourne Research Centre, Sittingbourne, Kent, U.K. 42. p.

Pandard, P., P. Vasseur, and D.M. Rawson, 1993. Comparison of two types of sensors using eukaryotic algae to monitor pollution of aquatic systems. *Water Res., 27:*427–431.

Parker, W.R. and N.A. Smith, 1997. A synopsis of the first cycle of the pulp and paper mill environmental effects monitoring program in the Atlantic Region. EPS-5-AR-99-2, Environment Canada, Environ. Prot. Branch, Atlantic Region, Surveillance, Halifax, N.S., 45 p.

Parkhurst, B.R., 1996. Predicting receiving system impacts from effluent toxicity. p. 309–321, In: *Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins (eds.), Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Parrott, J.L. and P.V. Hodson, 1995. Relative potency of polychlorinated dibenzo-*p*-dioxins and dibenzofurans for inducing mixed-function oxygenase activity in rainbow trout. *Environ. Toxicol. Chem., 14:*104--1050.

Payne, J.F., 1978. Mixed function oxidases in marine organisms in relation to petroleum hydrocarbon metabolism and detection. *Mar. Pollut. Bull., 8:*112–116.

Payne, J.F. and L.L. Fancey, 1982. Effect of long-term exposure to petroleum on mixed-function oxygenases in fish: further support for use of the enzyme system in biological monitoring. *Chemosphere, 11:*207–213.

Payne, J.F., L.L. Fancey, A.D. Rahimtula, and E.L. Porter, 1987. Review and perspective on the use of mixed-function oxygenase enzymes in biological monitoring. *Comp. Biochem. Physiol., 86C:*233–245.

Percy, J.A., 1977. Responses of arctic marine benthic crustaceans to sediments contaminated with crude oil. *Environ. Pollut., 13:*1–10.

Persaud, D., R. Jaagumagi, and A. Hayton, 1992. Guidelines for the protection and management of aquatic sediment quality in Ontario. PIBS 1962, Log 92-2309- 067, Ontario Min. of Environment, Water Resources Branch, Toronto, Ont., 26 p.

Persoone, G., 1998a. Development and first validation of a "stock-culture free" algal microbiotest: The Algaltoxkit. p. 311–320, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

———. 1998b. Development and validation of toxkit microbiotests with invertebrates, in particular crustaceans. p. 437–449, In: 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G., Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Persoone, G., A. Van de Vel, M. Van Steertegem, and B. De Nayer, 1989. Predictive value of laboratory tests with aquatic invertebrates: influence of experimental conditions. *Aquat. Toxicol., 14:*149–166.

Petersen, G.I. and P. Kristensen, 1998. Bioaccumulation of lipophilic substances in fish early life stages. *Environ. Toxicol. Chem.., 17:*1385–1395.

Peterson, H.G., F.P. Healey and R. Wagemann, 1984. Metal toxicity to algae: A highly pH dependent phenomenon. *Can. J. Fish. Aquat. Sci., 41:*974–979.

Pfleeger, T.G., A. Fong, R. Hayes, H. Ratsch, and C. Wickliff, 1996. Field evaluation of the EPA (Kenaga) nomogram, a method for estimating wildlife exposure to pesticide residues on plants. *Environ. Toxicol. Chem., 15:* 535–543.

Pickering, Q.H., 1988. Evaluation and comparison of two short-term fathead minnow tests for estimating chronic toxicity. *Water Res., 22:*883–893.

Pickering, Q.H., C. Henderson, and A.E. Lemke, 1962. The toxicity of organic phosphorus insecticides to different species of warmwater fishes. *Trans. Amer. Fish. Soc., 91:*75–184.

Pigeon, B., 1992. The Chemiotox index: a toxic waste indicator. Methodology applied to the 50 priority industrial plants under the St. Lawrence Action Plan. Prepared for L'Equipe d'intervention Saint-Laurent, Montréal, Québec, by BPC Environnement, 32 p.

Pillard, D.A., W.A. Stubblefield, and C. McPherson, 1993. An evaluation of sediment grain size as a confounding factor in assessing toxicity in shoreline sediment samples. *Fourteenth Ann. Meeting of Soc. Environ. Toxicol. Chem.*, Houston, Texas. Abstract 337.

Pocklington, P., and K.G. Doe, 1998. Development of a Canadian marine toxicity test for whole sediments using cultured spionid polychaetes. p. 395–407, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice*. P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Pollutech [Pollutech Enviroquatics Ltd.], 1996. Comparison of results from alternative acute toxicity tests with rainbow trout for selected mine effluents. Unnumbered AETE Report, April, 1996, Mining Association of Canada, and Natural Resources Canada, Canada Centre for Mineral and Energy Technol., Aquatic Effects Technol. Eval. Program, Ottawa, Ont.

Pollution Control Board, 1977 [reprinted 1989]. Pollution control objectives for the forest products industry of British Columbia. British Columbia Ministry of the Environment, Victoria, B.C.

Pontasch, K.W., B.R. Niederlehner, and J. Cairns Jr., 1989. Comparisons of single-species, microcosm and field responses to a complex effluent. *Environ. Toxicol. Chem., 8:*521–532.

Power, M. and L.S. McCarty, 1997. Fallacies in ecological risk assessment practices. *Environ. Sci. Technol./News 331:*370A–375A.

Prati, L., R. Pavanello, and F. Pesarin, 1971. Assessment of surface water quality by a single index of pollution. *Water Res., 5:*741–751.

Pratt, J.R., N.J. Bowers, and J. Cairns Jr., 1990. Effect of sediment on estimates of diquat toxicity in laboratory microcosms. *Water Res., 24:*51–57.

Quillardet, P., O. Huisman, R. D'Ari, and M. Hofnung, 1982. SOS Chromotest, a direct assay of induction of an SOS function in *Escherichia coli* K-12 to measure genotoxicity. *Proc. Nat. Acad. Sci. USA, 79:*5971–5975.

Quillardet, P., C. de Bellecombe, and M. Hofnung, 1985. The SOS Chromotest, a colorimetric bacterial assay for genotoxins: validation study with 83 compounds. *Mutat. Res., 147:*79–95.

Qureshi, A.D., K.W. Flood, S.R. Thompson, S.M. Junhurst, C.S. Inniss, and D.A. Rokosh, 1982. Comparison of a luminescent bacterial test with other bioassays for determining toxicity of pure compounds and complex effluents. p. 179–195, In: *Aquatic toxicology hazard assessment.* 5th conf. ASTM STP 766, J.G. Pearson *et al.* (eds.), Amer. Soc. for Testing and Materials, Philadelphia, Pa.

Qureshi, A.A., A.A.Bulich, and D.L. Isenberg, 1998. Microtox® toxicity test systems – where they stand today. p. 185–199, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* Wells, P.G., K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Ramamoorthy, S., 1996. *Handbook of chemical toxicity profiles of biological species. Vol. 2, Avian and mammalian species.* Lewis Pub., Boca Raton, Fla. 416 p.

Ramamoorthy, S. and E. Baddaloo, 1990. *Evaluation of environmental data for regulatory and impact assessment.* Elsevier, Amsterdam. 466 p.

———. 1995. *Handbook of chemical toxicity profiles of biological species. Vol. 1, Aquatic species.* Lewis Pub., Boca Raton, Fla. 400 p.

Rand, G.M. (ed.), 1995. *Fundamentals of aquatic toxicology. 2nd ed.* Taylor and Francis, Washington, D.C. 1125 p.

Rand, G.M. and S.R. Petrocelli (eds.), 1985. *Fundamentals of aquatic toxicology.* Hemisphere Pub., Washington, D.C.

Rawson, D.M., A.J. Willmer, and M.C. Cardosi, 1987. The development of a whole cell biosensor for on line screening of herbicide pollution of surface waters. *Toxic. Assess., 2:*325–340. [Cited through Pandard *et al.*, 1993].

Read, H.W., J.M Harkin, and K.E. Gustavson, 1998. Environmental applications with submitochondrial particles. p. 31–52, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Reice, S.R. and M. Wohlenberg, 1993. Monitoring freshwater benthic macroinvertebrates and benthic processes: measures for assessment of ecosystem health. p. *287–304, In: Freshwater biomonitoring and benthic macroinvertebrates.* D.M. Rosenberg and V.H. Resh (eds.), Chapman & Hall, New York, N.Y.

Reynoldson, T.B. and M.A. Zarull, 1993. An approach to the development of biological sediment guidelines. p. 177–200, In: *Ecological integrity and the management of ecosystems.* S. Woodley, J. Kay, and G. Francis (eds.), St. Lucie Press, Delray Beach, Fla.

Reynoldson, T.B., K.E. Day, and R.H. Norris, 1995. Biological guidelines for freshwater sediments based on

Benthic Assessment of SedimenT (the *BEAST*). *Austral. J. Ecol., 20:*198–219.

Roch, M. and E.J. Maly, 1979. Relationship of cadmium-induced hypocalcemia with mortality in rainbow trout *(Salmo gairdneri)* and the influence of temperature on toxicity. *J. Fish. Res. Board Can., 36:*1297–1303.

Rolland, R.M., M. Gilbertson, and R.E. Peterson (eds.), 1997. Chemically induced alteration in functional development and reproduction of fishes. Soc. Environ. Toxicol. Chem., SETAC Press, Pensacola, Fla., 224 p.

Rosenberg, D.M. and V.H. Resh (eds.), 1993. *Freshwater biomonitoring and benthic macroinvertebrates.* Chapman & Hall, New York, N.Y.

Rosenzweig, M.S. and A.L. Buikema Jr., 1994. Phytoplankton colonization and seasonal succession in new experimental ponds. *Environ. Toxicol. Chem., 13:*599–605.

Ross, P., 1998. Role of microbiotests in contaminated sediment assessment batteries. p. 549–556, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice*. P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Rousseaux, C.G., A. Branchaud, and P.A. Spear, 1995. Evaluation of liver histopathology and EROD activity in St. Lawrence Lake sturgeon *(Acipenser fulvescens)* in comparison with a reference population. *Environ. Toxicol. Chem., 14:*843–850.

Ruby, S.M., D.R. Idler, and S.Y. Peng, 1993. Plasma vitellogenin, 17β-estradiol, T_3 and T_4 levels in sexually maturing rainbow trout *Oncorhynchus mykiss* following sublethal HCN exposure. *Aquat. Toxicol., 26:*91–102.

Rue, W.J., J.A. Fava, and D.R. Grothe, 1988. A review of inter- and intralaboratory effluent toxicity test method variability. p. 190–203, In: Aquatic toxicology and hazard evaluation. 10th vol. W.J. Adams, G.A. Chapman, and W.G. Landis (eds.), Amer. Soc. Testing and Materials, Philadelphia, Pa., ASTM STP 971.

Russell, R.W., S.J. Hecnar, and G.D. Haffner 1995. Organochlorine pesticide residues in southern Ontario spring peepers. *Environ. Toxicol. Chem., 14:*815–818.

Russo, R.C., 1985. Ammonia, nitrite, and nitrate. p. 455–471, In: *Fundamentals of aquatic toxicology. Methods and applications.* G.M. Rand and S.E. Petrocelli (eds.), Hemisphere Pub. Corp., Washington, D.C.

Russom, C.L., S.P. Bradbury, S.J. Broderius, D.E. Hammermeister, and R.A. Drummond, 1997. Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow *(Pimephales promelas). Environ. Toxicol. Chem., 16:*948–967.

Salazar, M.H. and S.M. Salazar, 1996. Mussels as bioindicators: effects of TBT on survival, bioaccumulation, and growth under natural conditions. p. 305–330, In: *Organotin.* M.A. Champ and P.F. Seligman (eds.), Chapman & Hall, London.

Sanders, W.M., 1985. Field validation. p. 601–618, In: *Fundamentals of aquatic toxicology. Methods and applications.* G.M. Rand and S.E. Petrocelli (eds.), Hemisphere Pub. Corp., Washington, D.C.

Sarda, N. and G.A. Burton Jr., 1995. Ammonia variation in sediments: Spatial, temporal and method-related effects. *Environ. Toxicol. Chem., 14:*1499–1506.

Sasson-Brickson, G. and G.A. Burton Jr., 1991. In situ and laboratory sediment toxicity testing with *Ceriodaphnia dubia. Environ. Toxicol. Chem., 10:*201–207.

Sauvé, S., A. Dumestre, M. McBride, and W. Hendershot, 1998. Derivation of soil quality criteria using predicted chemical speciation of Pb^{2+} and Cu^{2+}. *Environ. Toxicol. Chem., 17:*1481–1489.

Schiewe, M.H., E.G. Hawk, D.I. Actor, and M.M. Krahn, 1985. Use of a bacterial luminescence assay to assay toxicity of contaminated marine sediments. *Can. J. Fish. Aquat. Sci., 42:*1244–1248.

Schimmel, S.C. and G.B. Thursby, 1996. Predicting receiving system impacts from effluent toxicity: a marine perspective. p. 322–330, In: *Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts.* Grothe, D.R., K.L. Dickson, and D.K. Reed-Judkins (eds.). Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Schimmel, S.C., G.E. Morrison, and M.A. Heber, 1989. Marine complex effluent toxicity program: test sensitivity, repeatability and relevance to receiving water toxicity. *Environ. Toxicol. Chem., 8:*739–746.

Schindler, D.W., 1987. Detecting ecosystem responses to anthropogenic stress. *Can. J. Fish. Aquat. Sci. 44 (Suppl. 1):*6–25.

Schrap, S.M. and A. Opperhuizen, 1990. Relationship between bioavailability and hydrophobicity: reduction of the uptake of organic chemicals by fish due to the sorption on particles. *Environ. Toxicol. Chem., 9:*715–724.

SCOPE (Scient. Committee on Problems of the Environment), 1975. Environmental impact assessment: principles and procedures. Internat. Counc. Scient. Unions, Scient. Committee on Problems of the Environ. (SCOPE), Paris. Scope Rept 5, 173 p.

———. 1980. Environmental risk assessment. (A.V. Whyte and I. Burton, eds.) Internat. Counc. Scient. Unions, Scient. Committee on Problems of the Environ. (SCOPE), Paris. Scope Rept 15, 157 p. John Wiley & Sons, New York.

Scroggins, R.P., 1986. In-plant toxicity balances for a bleached kraft pulp mill. *Pulp and Paper Canada, 87 (9):* T344–348.

Seelye, J.G., R.J. Hesselberg, and M.J. Mac, 1982. Accumulation by fish of contaminants released from dredged sediments. *Environ. Sci. Technol., 16:*459–464.

Sergy, G.A., 1987. Recommendations on aquatic biological tests and procedures for environmental protection, C & P DOE. Envirorment Canada, Environmental Protection, Technology Development and Technical Services Branch, Edmonton, Alberta. 102 p. + app.

Servos, M., J. Carey, M. Ferguson, G. Van Der Kraak, H. Ferguson, J. Parrott, K. Gorman, and R. Cowling, 1992. Impact of a modernized bleached kraft mill on white sucker populations in the Spanish River, Ontario. *Water Pollut. Res. J. Can., 27:*423–437.

SETAC (Soc. Environ. Toxicol. Chem.), 1997. Ecological risk assessment. Soc. Environ. Toxicol. Chem., Pensacola, Fla., SETAC Tech. Issue Paper, 4 p.

Sheppard, S., J. Bembridge, M. Holmstrup, and L. Posthuma, 1998. Advances in earthworm ecotoxicology. Soc. Environ. Toxicol. Chem., SETAC Press, Pensacola, Fla., 442 p.

Shi, M. and E.M. Faustman, 1989. Development and characterization of a morphological scoring system for medaka *(Oryzias latipes)* embryo culture. *Aquat. Toxicol., 15:*127–140.

Shigehisa, H. and H. Shiraishi, 1998. Biomonitoring with shrimp to detect seasonal change in river water toxicity. *Environ. Toxicol. Chem., 17:*687–694.

Shugart, L.R., 1995. Environmental genotoxicology. p. 405–419, In: *Fundamentals of aquatic toxicology. 2nd ed.* G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

Shugart, R., J. Bickham, G. Jackim, G. McMahon, W. Ridley, J. Stein, and S. Steinert, 1992. DNA alterations. p. 125–153, In: *Biomarkers. Biochemical, physiological, and histological markers of anthropogenic stress.* R.J. Huggett, R.A. Kimerle, P.M. Mehrle Jr., and H.L. Bergman (eds.), Lewis Pub., Boca Raton, Fla.

Sibley, P.K., G.T. Ankley, A.M. Cotter, and E.N. Leonard, 1996. Predicting chronic toxicity of sediments spiked with zinc: an evaluation of the acid-volatile sulfide model using a life-cycle test with the midge *Chironomus tentans. Environ. Toxicol. Chem., 15:*2102–2112.

Slooff, W., J.H. Canton, and J.L.M. Hermens, 1983. Comparison of the susceptibility of 22 freshwater species to 15 chemical compounds. I. (Sub)acute toxicity tests. *Aquat. Toxicol., 4:*113–128.

Slooff, W., J.A.M. van Oers, and D. de Zwart, 1986. Margins of uncertainty in ecotoxicological hazard assessment. *Environ. Toxicol. Chem., 5:*841–852.

Smith, S.L., D.D. MacDonald, K.A. Keenleyside, and C.L. Gaudet, 1996a. The development and implementation of Canadian sediment quality guidelines. p. 233–249, In: *Developments and progress in sediment quality assessment: Rationale, challenges, techniques and strategies.* M. Munawar, and G. Dave (eds.). Ecovision World Monograph Series. SPB Academic Publishing, Amsterdam, the Netherlands.

Smith, S.L., D.D. MacDonald, K.A. Keenleyside, C.G. Ingersoll, and L.J. Field, 1996b. A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. *J. Great Lakes Res., 22:*624–638.

Smrchek, S., R. Clements, R. Morcock, and W. Robert, 1993. Assessing biological hazard under TSCA: Methods and evaluation of data. p. 22–39, In: *Environmental toxicology and risk assessment.* W. Landis, J. Hughes, and M. Lewis (eds.), Amer. Soc. for Testing and Materials, Philadelphia, Pa., ASTM STP 1179.

Snell, T.W. and G. Persoone, 1989a. Acute toxicity bioassays using rotifers. I. A test for brackish and marine environments with *Brachionus plicatilis. Aquat. Toxicol., 14:*65–80.

Snell, T.W. and G. Persoone, 1989b. Acute toxicity bioassays using rotifers. II. A freshwater test with *Brachionus rubens. Aquat. Toxicol., 14:*81–92.

SNV (Swedish Naturvårdsverket), 1990. Biological – chemical characterisation of industrial waste water. Application when granting permits and exercising supervisory authority for activities harmful to the environment. Naturvårdsverket (Swedish Environmental Protection Agency), S-171 85 Solna, Sweden. 102 p. [ISBN 91-620-1071-9]

Soares, A.M.V.M., D.J. Baird, and P. Calow, 1992. Interclonal variation in the performance of *Daphnia magna* Straus in chronic bioassays. *Environ. Toxicol. Chem., 11:*1477–1484.

Södergren, A., 1987. Biological effects of effluents from pulp mills – preliminary results from the Swedish Environment/Cellulose project. *Paperi ja Puu, 69:*422–426.

Södergren, A., B.-E. Bengtsson, P. Jonsson, S. Lagergren, L. Larsson, M. Olsson and L. Renberg, 1988. Summary of results from the Swedish project Environment/Cellulose. *Water Sci. Technol., 20:*49–60.

Solé, M., C. Porte, and J. Albaigés, 1995. Seasonal variation in the mixed-function oxygenase system and antioxidant enzymes of the mussel *Mytilus galloprovincialis*. *Environ. Toxicol. Chem., 14:*157–164.

Spear, P.A. and R.C. Pierce, 1979. Copper in the aquatic environment: chemistry, distribution and toxicology. National Research Council of Canada, Environ. Secretariat, Rept no. NRCC 16454.

Spies, R.B. and D.W. Rice Jr., 1988. Effects of organic contaminants on reproduction of the starry flounder *Platichthys stellatus* in San Francisco Bay. II. Reproductive success of fish captures in San Francisco Bay and spawned in the laboratory. *Marine Biol., 98:*191–200.

Sprague, J.B., 1969. Measurement of pollutant toxicity to fish – I. Bioassay methods for acute toxicity. *Water Res. 3:*793–821.

———. 1970. Measurement of pollutant toxicity to fish. II. Utilizing and applying bioassay results. *Water Res., 4:* 3–32.

———1971. Measurement of pollutant toxicity to fish – III. Sublethal effects and "safe" concentrations. *Water Res., 5:*245–266.

———. 1981. Ethologists are environmental drop-outs. Abstract. Proc. 2nd Biennial Symposium on Ethology and Behavioural Ecology of Fishes, Normal, Ill., Oct. 19-22, 1979. *Environ. Biol. Fishes, 6:*138.

———. 1985. Factors that modify toxicity. p. 124–163, In: *Fundamentals of aquatic toxicology. Methods and applications.* G.M. Rand and S.E. Petrocelli (eds.), Hemisphere Pub. Corp., Washington, D.C.

———. 1991. Environmentally desirable approaches for regulating effluents from pulp mills. *Water Sci. Technol,. 3/4:*361–271.

———. 1992. Concepts for developing Canadian water quality guidelines for effluents from bleached pulp mills. Prepared for Environment Canada, Ecosystem Sciences and Evaluation Directorate, Eco-Health Branch, Guidelines Div., Hull, Quebec. 129 p.

———. 1995. Planning a field validation program. Chapt. 3 in Review of methods for sublethal aquatic toxicity tests relevant to the Canadian metal-mining industry, and design of field validation programs. Prepared for Natural Resources Canada, Canada Centre for Mineral and Energy Technology, Aquatic Effects Technol. Eval. Progr. [AETE], Ottawa, Ont. by Sprague Associates Ltd., Salt Spring Island, B.C., 110 p.

———. 1997. Review of methods for sublethal aquatic toxicity tests relevant to the Canadian metal-mining industry. Natural Resources Canada, Canada Centre for Mineral and Energy Technology, Aquatic Effects Technol. Eval. Progr. [AETE], Ottawa, Ont. AETE Project 1.2.1: 102 p.

Sprague, J.B. and D.E. Drury, 1969. Avoidance reactions of salmonid fish to representative pollutants. p. 169–179 In: *Advances in Water Pollution Research. Proc. Fourth Internat. Conf., Prague, Vol. 1.* S.H. Jenkins (ed.), Pergamon Press, London.

Sprague, J.B., P.F. Elson, and R.L. Saunders, 1965. Sublethal copper-zinc pollution in a salmon river – a field and laboratory study. *Internat. J. Air Water Pollution, 9:*531–543.

Sprague, J.B., P.F. Elson, and J.R. Duffy, 1971. Decrease in DDT residues in young salmon after forest spraying in New Brunswick. *Environ. Pollut., 1:*191–203.

Sprague, J.B., N. McCubbin, and N. Bonsor [Expert Committee on Pollution from Non-Kraft Pulp and Paper Mills], 1991. Effluents from non-kraft pulp and paper mills in Ontario. Report for Technical Advisory Committee, Pulp and Paper Sector of MISA, Ontario Min. of Environment, Toronto, Ont., 426 p.

SRC (Saskatchewan Research Council) 1995a. Development of aquatic plant bioassays for rapid screening and interpretive risk assessments of metal mining wastewaters. SRC, Tech. Univ. Denmark, and Univ. Saskatchewan, Regina, Sask. SRC Pub. E-2100-2-C-95, 191 p.

Stackhouse, R.A. and W.H. Benson, 1988. The influence of humic acid on the toxicity and bioavailability of selected trace metals. *Aquatic Toxicol., 13:*99–108.

Stebbing, A.R.D., 1982. Hormesis – the stimulation of growth by low levels of inhibitors. *Sci. Total Envir., 22:* 213–234.

Stegeman, J.J. and P.J. Kloepper-Sams, 1987. Cytochrome P-450 isozymes and mono-oxygenase activity in aquatic animals. *Environ. Health Perspect., 71:*87–95.

Stegeman, J.J., B.R. Woodin, and A. Boksoyr, 1988. Apparent cytochrome P-450 induction as an indication of exposure to environmental chemicals in the flounder *Platichthys flesus. Mar. Ecol. Prog. Ser., 46:*55–60.

Stein, J.E., T.K. Collier, W.L. Reichert, E. Casillas, T. Hom, and U. Varanasi, 1992. Bioindicators of contaminant exposure and sublethal effects: Studies with benthic fish in Puget Sound, Washington. *Environ. Toxicol. Chem., 11:* 701–714.

Stephan, C.E., 1982. Increasing the usefulness of acute toxicity tests. p. 69–81, In: *Aquatic toxicology and hazard assessment: fifth conference.* J.G. Pearson, R.B. Foster, and W.E. Bishop (eds.), Amer. Soc. Testing and Materials, Philadelphia, Pa. ASTM STP No. 766.

Stephan, C.E. and J.W. Rogers, 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. p. 328–338, In: *Aquatic toxicology and hazard assessment. 8th symp.* R.C. Bahner and D.J. Hansen (eds.), American Soc. Testing and Materials, Philadelphia, Pa., ASTM STP 891.

Stephan, C.E., D.I. Mount, D.J. Hanson, J.H. Gentile, G.A. Chapman, and W.A. Brungs, 1985. Guidelines for deriving numeric national water quality criteria for the protection of aquatic organisms and their uses. USEPA, Duluth, Minn. Natl Tech. Inform. Service, PB85-227049.

Stumm, W. and J.J. Morgan, 1981. *Aquatic chemistry. An introduction emphasizing chemical equilibria in natural waters. Second ed.* John Wiley & Sons, New York, N.Y.

Suter, G.W. II, 1990. Uncertainty in environmental risk assessment. p. 203–230, In: *Acting under uncertainty: Multidisciplinary conceptions.* G.M. von Furstenberg (ed.), Kluwer Academic Pub., Boston. Mass.

———. (ed.), 1993. *Ecological risk assessment.* Lewis Pub., Boca Raton, Fla. 538 p.

———. 1995. Introduction to ecological risk assessment for aquatic toxic effects. p. 803–816, In: *Fundamentals of aquatic toxicology. 2nd ed.* G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

———. 1996. Abuse of hypothesis testing statistics in ecological risk assessment. *Human and Ecol. Risk Assess., 2:*331–347.

———. 1997. Overview of the ecological risk assessment framework. p. 1–6, In: *Ecological risk assessment of contaminated sediments.* C.G. Ingersoll, T. Dillon, and G.R. Biddinger (eds.). Soc. Environ. Toxicol. Chem., SETAC Special Pub. Series, SETAC Press, Pensacola, Fla.

Suter, G.W. II, A.E. Rosen, E. Linder, and D.F. Parkhurst, 1987. Endpoints for responses of fish to chronic toxic exposures. *Environ. Toxicol. Chem., 6:*793–809.

Swanson, M.B., G.A. Davis, L.E. Kincaid, T.W. Schultz, J.E. Bartmess, S.L. Jones, and E.L. George, 1997. A screening method for ranking and scoring chemicals by potential human health and environmental impacts. *Environ. Toxicol. Chem., 16:*372–383.

Tang, Y. and E.T. Garside, 1987. Preexposure and subsequent resistance to lead in yearling brook trout, *Salvelinus fontinalis. Can. J. Fish. Aquat. Sci., 44:*1089–1091.

Taub, F.B., 1985. Toward interlaboratory (round-robin) testing of a standardized aquatic microcosm. p. 165–186, In: *Multispecies toxicity testing.* Cairns, J. Jr. (ed.), Pergamon Press, New York, N.Y.

Tay, K.-L., K.G. Doe, S.J. Wade, D.A. Vaughan, R.E. Berrigan, and M.J. Moore, 1992. Sediment bioassessment in Halifax harbour. *Environ. Toxicol. Chem., 11:*1567–1582.

Tay, K.L., K.G. Doe, A.J. MacDonald, S.J. Wade, J.D.A. Vaughan, A.L. Huybers, K. Lee, and R. Larocque, 1993. Application of sediment bioassays for the environmental assessment of ocean dumping sites. *Proc. 20th Ann. Aquatic Toxicity Workshop,* Quebec (Quebec), November, 1993. p. 48-49.

Tay, K.L., K.G. Doe, A.J. MacDonald, and K. Lee, 1998. The influence of particle size, ammonia, and sulfide on toxicity of dredged materials for ocean disposal. p. 559–574, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Taylor, B.R. and D. Parkinson, 1988. A new microcosm approach to litter decomposition studies. *Can. J. Bot., 66:*1933–1939.

Taylor, M.C. (ed.), 1989. *Proceedings of a workshop on the development and use of water quality objectives.* Halifax, N.S., Sept. 19-21, 1989. Canadian Council of Ministers of the Environment, Task Force for Water Quality Guidelines, 175 p.

Thompson, K.W., M.L. Deaton, R.V. Foutz, J. Cairns Jr., and A.C. Hendricks, 1982. Application of time-series intervention analysis to fish ventilatory response data. *Can. J. Fish. Aquat. Sci., 39:*518–521.

Thursby, G.B. and C.E. Schlekat, 1993. Statistical analysis of 10-day solid phase toxicity data for amphipods. Fourteenth Ann. Meeting of Soc. Environ. Toxicol. Chem., Houston, Texas. Abstract 339.

Thurston, R.V., G.R. Phillips, R.C. Russo, and S.M. Hinkins, 1981. Increased toxicity of ammonia to rainbow trout *(Salmo gairdneri)* resulting from reduced concentrations of dissolved oxygen. *Can. J. Fish. Aquat. Sci., 38:*983–988.

Thurston, R.V., T.A. Gilfoil, E.L. Meyn, R.K. Zajdel, T.I. Aoki, and G.D. Veith, 1985. Comparative toxicity of ten organic chemicals to ten common aquatic species. *Water Res., 9:*1145–1155.

Tosato, M.L., S. Marchini, L. Passerini, A. Pino, L. Eriksson, F. Lindgren, S. Hellberg, J. Jonsson, M. Sjöström, B. Skagerberg, and S. Wold, 1990. QSARs based on statistical design and their use for identifying chemicals for further biological testing. *Environ. Toxicol. Chem., 9:*265–277.

Trust, K.A., A. Fairbrother, and M.J. Hooper, 1994. Effects of 7, 12-dimethylbenz[*a*]anthracene on immune function and mixed-function oxygenase activity in the European starling. *Environ. Toxicol. Chem., 13:*821–830.

USEPA (United States Environmental Protection Agency), 1978. Water quality criteria. *U.S. Fed. Register, 43(97):*21506–21518.

———. 1984a. Development of water quality-based permit limitations for toxic pollutants; national policy. U.S. *Federal Register, 49(*48)*:* 9016–9019, Friday, March 9, 1984.

———. 1984b. Estimating "concern levels" for concentrations of chemical substances in the environment. USEPA, Environmental Effects Branch, Health and Environmental Review Div., 43 p.

———. 1985a. Ambient aquatic life water quality criteria for ammonia – 1984. EPA/440/5-85-001. USEPA, Office of Water, Regulations and Stand., Washington, D.C.

———. 1985b. Toxic Substances Control Act test guidelines; final rules. *U.S. Federal Register, 50:*39252–39516, Sept. 27, 1985.

———. 1986. Quality criteria for water 1986. EPA 440/5-86-001. USEPA, Office of Water, Regulations and Standards, Washington, D.C.

———. 1988a. Toxicity reduction evaluation protocol for municipal wastewater treatment plants. EPA/600/2-88/062. USEPA, Office of Research and Development, Risk Reduction Engineering Lab., Cincinnati, Ohio.

———. 1988b. Generalized methodology for conducting industrial toxicity reduction evaluations. EPA/600/2-88/070. USEPA, Office of Research and Development, Risk Reduction Engineering Lab., Cincinnati, Ohio.

———. 1989a. Methods for aquatic toxicity identification evaluations. Phase II. Toxicity identification procedures. EPA/600/3-88/035. USEPA, Office of Research and Development, Environmental Research Laboratory, Duluth, Minn.

———. 1989b. Methods for aquatic toxicity identification evaluations. Phase III. Toxicity confirmation procedures. EPA/600/3-88/036. USEPA, Office of Research and Development, Environmental Research Laboratory, Duluth, Minn.

———. 1989c. Ecological assessment of hazardous waste sites: a field and laboratory reference document. W. Warren-Hicks, B.R. Parkhurst, and S.S. Baker (eds.), Prepared by Kilkelly Environmental Associates. EPA/600/3-89/013. USEPA,

———. 1989d. Federal Insecticide, Fungicide and Rodenticide Act (FIFRA): good laboratory practice standards; final rule. 40 CFR Part 160. *U.S. Federal Register, IV,* August 17, 1989. 34 p.

———. 1989e. Amended good laboratory practices (TSCA). *U.S. Federal Register*, August 17, 1989.

———. 1991a. Technical support document for water quality-based toxics control. EPA/505/2-90-001. USEPA, Office of Water, Washington, D.C.

———. 1991b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to marine and estuarine organisms. 2nd ed. EPA 600/4-91-003. USEPA, Environmental Monitoring Systems Laboratory -- Cincinnati, Ohio.

———. 1991c. Methods for aquatic toxicity identification evaluations. Phase I. Toxicity characterization procedures. EPA/600/6-91/003. USEPA, Office of Research and Development, Environmental Research Laboratory, Duluth, Minn.

———. 1991d. Toxicity identification evaluation: characterization of chronically toxic effluents, phase I. EPA/600/6-91/005. USEPA, Office of Research and Development, Environmental Research Laboratory, Duluth, Minn.

———. 1992a. Application of microcosms for assessing the risk of microbial biotechnology products. In:EPA workshop report. EPA/600/R-92/066. C.R. Cripe, P.H. Pritchard, and A.M. Stern (eds.), USEPA, Washington, D.C.

———. 1992b. Framework for ecological risk assessment. Risk Assessment Forum. EPA/630/R-92/001. USEPA, Washington, D.C.

———. 1992c. Sediment classification methods compendium. EPA 823-R-92-006. USEPA, Office of Water, Washington, D.C.

———. 1993. Guidance manual: bedded sediment bioaccumulation tests. EPA/600/R-93/183. USEPA, Office of Research and Development, Washington, D.C.

———. 1994a. Short-term methods for estimating the chronic toxicity of effluent and receiving waters to marine and estuarine organisms. 2nd. ed. EPA/600/4-91/003., USEPA, Cincinnati, Ohio.

———. 1994b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. 3rd ed. EPA 600/4-91-002. USEPA, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio.

———. 1994c. ECOSAR. A computer program for estimating the ecotoxicity of industrial chemicals based on structure activity relationships. User's guide. Rept no. 748-R-93-002, USEPA, Office Pollut. Prevention and Toxics, Washington, D.C., 22 p. and computer disc.

———. 1994d. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. EPA 600/R-94/024, USEPA, Duluth, Minn., 32 p.

———. 1994e. Water quality standards handbook: second edition. EPA/823-B-94-005a. USEPA, Office of Water, Washington, D.C.

———. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. EPA 600/R-95/136, G.A. Chapman, D.L. Denton, and J.M. Lazorchak (eds.), USEPA, Office of Research and Development, Washington, D.C. 661 p.

———. 1996. Marine toxicity identification evaluation (TIE): Phase I, guidance document. EPA/600/R-96/054. USEPA, Cincinnati, Ohio.

———. 1998. Guidelines for ecological risk assessment. EPA/630/R-95/002Fa. USEPA, Risk Assessment Forum, Washington, D.C. [Available as NTIS pub. no. PB98-117849 or from http://www.epa.gov/ncea/]

USEPA Region VIII (United States Environmental Protection Agency), 1994. Mixing zones and dilution policy. USEPA Region VIII, Water Management Div., Denver, Colo. 38 p + app.

USEPA and ACOE (United States Environmental Protection Agency and U.S. Army Corps of Engineers), 1994. Evaluation of dredged material proposed for discharge in waters of the U.S. USEPA, Offices of Water and Science & Technology, and U.S. Army Corps of Engineers, Washington, D.C., 161 p. + app.

USEPA and EEC (United States Environmental Protection Agency and European Economic Community), 1993. USEPA/EEC joint project on the evaluation of (quantitative) structure-activity relationships. (Draft final report). USEPA, Washington, D.C., 55 p. + app.

USNRC (United States National Research Council), 1989. Improving risk communication. National Academy Press, Washington, D.C., 332 p.

Vaes, W.H.J., E.U. Ramos, H.J.M. Verhaar, and J.L.M. Hermens, 1998. Acute toxicity of nonpolar versus polar narcosis: Is there a difference? *Environ. Toxicol. Chem., 17:*1380–1384.

Valiela, D. and P.H. Whitfield, 1988. Designing site-specific water quality objectives and monitoring. *Water Pollut. Res. J. Can., 23:*510–519.

van de Plassche, E.J., and G.J. Bockting, 1993. Towards integrated environmental quality objectives for several volatile compounds. Report no. 679101-011, National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands, 78 p.

van der Hoeven, N. and A.A.M. Gerritsen, 1997. Effects of Chlorpyrifos on individuals and populations of *Daphnia pulex* in the laboratory and field. *Environ. Toxicol. Chem., 16:*2438–2447.

van der Schalie, W.H., 1986. Can biological monitoring early warning systems be useful in detecting toxic materials in water? p. 107–121, In: Aquatic toxicology and risk assessment. 9th vol., ASTM STP 921. T.M. Poston and R. Purdy (eds.), Amer. Soc. for Testing and Materials, Philadelphia, Pa.

Van Gestel, C.A.M. and W.- C. Ma, 1993. Development of QSAR's in soil ecotoxicology: Earthworm toxicity and soil sorption of chlorophenols, chlorobenzenes and chloroanilines. *Water Air Soil Pollut., 69:*265–276.

van Hoof, F.M., E.G. de Johghe, M.G. Briers, P.D. Hansen, H.J. Pluta, D.M. Rawson, and A.J. Wilmer, 1992. The evaluation of bacterial biosensors for screening of water pollutants. *Environ. Toxicol. Water Qual., 7:*19–33.

Van Loon, W.M.G.M., M.E. Verwoerd, F.G. Wijnker, C.J. van Leeuwen, P. van Duyn, C. Van de Guchte, and J.L.M. Hermens, 1997. Estimating total body residues and baseline toxicity of complex organic mixtures in effluents and surface waters. *Environ. Toxicol. Chem., 16:*1358–1366.

Van Sprang, P.A., and C.R. Janssen, 1997. Identification and confirmation of ammonia toxicity in contaminated sediments using a modified toxicity identification evaluation approach. *Environ. Toxicol. Chem., 16:*2501–2507.

van Straalen, N.M. and C.A.J. Denneman, 1991. Ecotoxicological evaluation of soil quality criteria. *Ecotox. Environ. Saf., 18:*241–251.

Van Urk, G., F. Kerkum, and C.J. Van Leeuwen, 1993. Insects and insecticides in the Lower Rhine. *Water Res., 27:*205–213.

Van Voris, P., D.A. Tolle, M.F. Arthur, and J. Chesson, 1985. Terrestrial microcosms: Applications, validation and cost-benefit analysis. p. 117–142, In: *Multispecies toxicity testing*. J. Cairns Jr. (ed.), Pergamon Press, New York, N.Y.

Veith, G.D., D.L. DeFoe, and B.V. Bergstedt, 1979. Measuring and estimating the bioconcentration factor of chemicals in fish. *J. Fish. Res. Board. Can., 36:*1040–1048.

Veith, G.D., D.J. Call, and L.T. Brooke, 1983. Structure - toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. *Can. J. Fish. Aquat. Sci., 40:*743–748.

Verhaar, H.J.M., C.J. van Leewen, and J.L.M. Hermens, 1992. Classifying environmental pollutants. 1: structure-activity relationships for prediction of aquatic toxicity. *Chemosphere, 25:*471–491.

VKI (Water Quality Institute), 1994. Discussion paper regarding guidance for terrestrial effects assessment. Draft final report, prepared for the Organisation for Economic Cooperation and Development. VKI, Water Quality Institute, Denmark. 63 p.

VKI and RIVM (Water Quality Institute and Institute of Public Health and Environmental Protection), 1995. Aquatic testing methods for pesticides and industrial chemicals. Review. Annex, draft final report, prepared for National Co-ordination of the OECD Test Guidelines Programme, Paris. VKI, Water Quality Institute, Denmark, and Institute of Public Health and Environmental Protection, the Netherlands, 332 p.

Voyer, R.A., P.O. Yevich, and C.A. Barszcz, 1975. Histological and toxicological responses of the mummichog *Fundulus heteroclitus* (L.) to combinations of levels of cadmium and dissolved oxygen in a freshwater. *Water Res., 9:*1069–1074.

Wagner, C. and H. Lϕkke, 1991. Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Res., 25:*1237–1242.

Wajsbrot, N., A. Gasith, M.D. Krom, and T.M. Samocha, 1990. Effect of dissolved oxygen and the molt stage on the acute toxicity of ammonia to juvenile green tiger prawn *Penaeus semisulcatus. Environ. Toxicol. Chem,. 9:*497–504.

Walker, G.C., 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *E. coli. Microb. Rev., 48:*60–93.

Walker, J.D., 1990. Chemical fate, bioconcentration, and environmental effects testing: proposed testing and decision criteria. *Tox. Assess., 5:*103–134.

Wall, T.M. and R.W. Hanmer, 1987. Biological testing to control toxic water pollutants. *J. Water Pollut. Control Fed., 59:*7–12.

Wall, V.D., J. London, J.E. Warren, R. Gossett, M.D. Wenholz, and S.J. Klaine, 1998. Development of a continuous-flow renewal system for sediment toxicity testing. *Environ. Toxicol. Chem., 17:*1159–1164.

Wallace, J.B., T.F. Cuffney, J.R. Webster, G.J. Lugthart, K. Chung, and B.S. Goldowitz, 1991. Export of fine organic particles from headwater streams: Effects of season, extreme discharges, and invertebrate manipulation. *Limnol. Oceanogr., 36:*670–682.

Wang, F., 1998. Porewater toxicity testing: does it make sense? Soc. Environ. Toxicol. Chemistry, Pensacola, Fla. *SETAC News, 18 (5):*23–24

Warren, C.E. and G.E. Davis, 1971. Laboratory stream research: Objectives, possibilities, and constraints. *Annual Rev. Ecol. Systematics, 2:*111–144.

Warren-Hicks, W.J., and D.R.J. Moore (eds), 1998. Uncertainty analysis in ecological risk assessment. Soc. Environ. Toxicol. Chem., SETAC Press, Pensacola, Fla., 314 p.

WSDOE (Washington State Department of Ecology), 1997. Laboratory guidance and whole effluent toxicity test review criteria. Pub. no. WQ-R-95-80, revised March 1997, WSDOE, Water Quality Program, Olympia, Wash., 62 p.

———. 1998. Whole effluent toxicity (WET) program evaluation. Pub. no. 98-03, WSDOE, Water Quality Program, Olympia, Wash., 31 p.

Wells, P.G., K. Lee, and C. Blaise (eds), 1998. *Microscale testing in aquatic toxicology. Advances, techniques, and practice.* CRC Press, Boca Raton, Fla.

Westlake, G.F., J.B. Sprague, and I.T. Brown, 1983. Sublethal effects of treated liquid effluent from a petroleum refinery. IV. Respiratory movements and coughing of rainbow trout. *Aquat. Toxicology, 4:*317–325.

Wexler, P (ed.), 1998. *Encyclopedia of toxicology.* Academic Press, New York, N.Y.

White, P.A., and C. Côté, 1998. Investigating the sources and fate of genotoxic substances in aquatic ecosystems with the SOS Chromotest. p. 607–630, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Whitehouse, P. and M. Mallet, 1993. Aquatic toxicity testing for notification of new substances. An advisory document on dealing with 'difficult' substances. U.K. Dept. of the Environment, Chemical Notification Unit, Rept No. CP 722, 34 p. [Publications Unit, P.O. Box 16, Marlow, Buckinghamshire, U.K.]

Wilson, R.C.H., R.H. Cook, and E. Pessah, 1975. Practical application of the toxicity emission rate concept in a survey of Maritime pulp and paper mills. Presented at *Can. Pulp and Paper Assoc. Tech. Section Conf.*, Sept. 1975, Vancouver, B.C., p. 33–38.

Winner, R.W., 1981. A comparison of body length, brood size and longevity as indices of chronic copper and zinc stresses in *Daphnia magna*. *Environ. Pollut., (A)26:*33–37.

Wren, C.D., G.L. Stephenson, and R.P. Lanno, 1993. Using earthworms during biological testing for soil quality. *Environ. Science. Engng.*, September 1993, 26–27.

Wright, J.F., D. Moss, P.D. Armitage, and M.T. Furse, 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. *Freshwat. Biol., 14:*221–256.

Xu, H.H., 1998. ATP-TOX system – a review. p. 153–159, In: *1998. Microscale testing in aquatic toxicology. Advances, techniques, and practice.* P.G. Wells, K. Lee, and C. Blaise (eds.), CRC Press, Boca Raton, Fla.

Xu, H. and B.R. Dutka, 1987. ATP-TOX system – a rapid, sensitive bacterial toxicity screening system based on the determination of ATP. *Toxic. Assess., 2:*149–166. [Cited through Dutka *et al.*, 1988].

Zeeman, M.G., 1995. Ecotoxicity testing and estimation methods developed under section 5 of the Toxic Substances Control Act (TSCA). p. 703–715, In: *Fundamentals of aquatic toxicology. 2nd ed.* G.M. Rand (ed.), Taylor and Francis, Washington, D.C.

Zischke, J.A., J.W. Arthur, R.O. Hermanutz, S.F. Hedtke, and J.C. Helgen, 1983. Effects of pentachlorophenol on invertebrates and fish in outdoor experimental channels. *Aquatic Toxicol., 7:*37–58.

Members of the Project Steering Committee, Development of a Guidance Manual on Environmental Toxicological Data Interpretation and Application
(as of September, 1997)

Robert Baker
Environmental Assessment Branch
Environmental Conservation Service
Environment Canada
Hull, Quebec

Norman Bermingham
Centre St. Laurent
Environment Canada
Montreal, Quebec

Philippa M. Cureton
Commercial Chemicals Evaluation Branch
Environmental Protection Service
Environment Canada
Hull, Quebec

Marie-Louise Geadah
National Office of Pollution Prevention
Environmental Protection Service
Environment Canada
Hull, Quebec

Karen Keenleyside
Evaluation and Interpretation Branch
Environmental Conservation Service
Environment Canada
Hull, (Quebec)

Ian MacKenzie
Standards and Approvals Division
Alberta Environmental Protection
Edmonton, Alberta

Jennifer A. Miller
Method Development and Application Section
Environmental Technology Centre
Gloucester, Ontario

James M. Osborne
Marine Environmental Division
Hazardous Waste Branch
Environmental Protection Service
Environment Canada
Hull, Quebec

W. Roy Parker
Atlantic Region
Environmental Protection Service
Environment Canada
Dartmouth, Nova Scotia

Richard P. Scroggins
Method Development and Application Section
Environmental Technology Centre
Gloucester, Ontario

Rob Stevens
Phys/Chem. Sciences Directorate
Fisheries and Oceans Canada
Ottawa, Ontario

Roger Sutcliffe
New Substance Division
Commercial Chemicals Evaluation Branch
Environmental Protection Service
Environment Canada
Hull, Quebec

The PEEP and CHIMIOTOX Tests Used in the Effluent Testing Program of the St. Lawrence Action Plan

The St. Lawrence Action Plan was a five-year, federal-provincial program aimed at cleaning up, restoring, protecting, and preserving the St. Lawrence River. The St. Lawrence flows through the most heavily industrialized region of Quebec, where a multitude of industries discharge effluent. These effluents vary widely in volume, chemical composition, and toxicity. One of the major objectives of the Plan was to reduce the liquid toxic waste by 90%, from the 50 major industrial plants along the St. Lawrence and Saguenay Rivers by 1993 (Costan *et al.*, 1993).

The first phase of the Plan intended to: (1) identify the main sources of pollution; (2) reduce toxic emissions from key industries; and (3) establish biological and chemical standards of effluent quality. To do this, a tool was needed to compare the toxicity and chemical composition of the 50 complex industrial effluents, each with over 100 potentially toxic constituents. Ideally, the tool should:

(1) compare and integrate the results of chemical analyses;
(2) evaluate conditions according to (a) industrial sectors (e.g., pulp and paper mills, metallurgical industries), (b) groups of industrial plants, (c) type of pollutant, and (d) overall comparison of the 50 industries; and
(3) assess the reductions in toxic discharges as industries improved.

Two tools were devised for measurements: the PEEP index to measure toxicity of effluent; and the Chimiotox index to rate the predicted toxic potential of an effluent from its chemical composition. These tools were developed for a particular region and their suitability for another location would deserve examination. Investigators should also keep in mind that any toxicity index consolidates several pieces of information with a loss of detailed information. In addition, the Chimiotox is based on chemistry and predicts but does not measure toxicity.

Potential Ecological Effects Probe (PEEP). The PEEP index was developed by the L'Equipe d'intervention Saint-Laurent, to rank the 50 major industrial effluents in terms of their potential impact on the river. The PEEP index was intended as a comprehensive, cost-effective, and sensitive index of effluent toxicity. The ranking by PEEP had two main purposes:

(1) to focus limited resources on the effluents with the greatest effects on the river; and
(2) to allow efficient monitoring of effluents so that improvements could be followed.

PEEP combines results of four small-volume toxicity tests. It appears to have worked well in assessing effluents on the St. Lawrence River (Legault and Villeneuve, 1992; Costan *et al.*, 1993).

The index is based on results of the following four small-volume screening tests, selected to incorporate (1) lethal and sublethal toxicity, (2) genetic toxicity, and (3) a range of trophic levels.

(1) Algal growth test with *Selenastrum capricornutum* (EC, 1992d)
(2) SOS-Chromotest with *Escherichia coli* (Quillardet *et al.*, 1982; 1985; Section 6.1)
(3) Luminescence test with *Vibrio fischeri* (formerly *Photobacterium phosphoreum)* (EC, 1992c)
(4) Survival and reproduction test with *Ceriodaphnia dubia* (EC, 1992a).

To apply the PEEP index, the effluent is first evaluated using all four toxicity tests.

One litre of effluent is then subjected to a process designed to simulate aerobic biological treatment. It includes incubation in order to measure the persistence of toxicity through mild microbial degradation, although volatile constituents will also be lost (Costan *et al.*, 1993). Biodegradation proceeds as follows: the litre of effluent is enriched with (1) a concentrated nutrient solution containing $MgSO_4$, $CaCl_2$, and $FeCl_2$, and (2) a phosphate buffer solution, pH 7.2, which also contains NH_4Cl. The sample is inoculated with a 5% commercial mixture of microbial seed and incubated for 5 d in darkness, at room temperature, with constant gentle aeration.

The sample is re-tested after incubation using three of the four toxicity tests, excluding the more expensive *Ceriodaphnia* test.

Results of all tests are reported as threshold concentrations for observed effect (TOEC), so a value can be entered into the PEEP index even if no ICp or EC50 was determined. The TOECs are converted to toxic units (TU = 100/TOEC, see Section 4.3.1). The PEEP is calculated as the sum of the TU, including both before and after biodegradation, according to the following formula.

$$P = \log_{10}\left[1 + n\left(\frac{\sum_{i-1}^{N} Ti}{N}\right) Q\right]$$

where: P = PEEP value
n = number of endpoints exhibiting toxic responses
N = maximum number of obtainable toxic endpoints
T_i = Toxic Units from each test, before or after biodegradation
Q = effluent flow in m^3/h

The four tests produce a total of six endpoints, because the *Ceriodaphnia* and SOS-Chromotests each have two endpoints. Without *Ceriodaphnia*, there are four endpoints after biodegradation, so the value of N is 10. The PEEP formula sums the TU from each endpoint and divides by N to derive an "average Toxic Units" for the effluent. That is multiplied by n to account for the breadth of toxic response, and by discharge to compute the total hourly number of toxic units entering the river (≈ TEF, see Section 4.3.3). The result is expressed as a logarithm (adding 1 in the formula merely ensures that the log is computable). This scales results so that they normally fall between 0 and 10 (Costan *et al.*, 1993). Thus, PEEP collapses results of 7 toxicity tests into a single number, with loss of the detailed information from individual tests, to acquire a single index.

The approach is designed to be flexible, and one or more tests can be removed or replaced with another test without preventing calculation of a value for PEEP. (Indices based on different tests would not be strictly comparable.) PEEP values from different industries are comparable since they are all based on toxicity. They allow for volume of discharge, as appropriate for considering environmental disturbance. The companion index, Chimiotox, is based on chemical measurements, so it is complementary.

Chimiotox. Chimiotox provides a physicochemical characterization of substances in an effluent. The substances are weighted for their known toxicity, with the goal of providing a common denominator in terms of potential environmental damage. The *toxic weighting factor* (F_{tox}) is calculated according to the following formula (Pigeon, 1992).

$$F_{tox_i} = \frac{1 mg/L}{MSC_i mg/L}$$

where: $F_{tox\ i}$ = toxic weighting factor for substance i
1 mg/L = arbitrary reference
MSC_i = most stringent water quality criterion (guideline) for substance i

Chimiotox units (UC) for each substance are calculated by multiplying its toxic weighting factor by its polluting load.

$$UC_i = Load_i \times F_{tox_i}$$

where: UC_i = Chimiotox unit of substance i
$Load_i$ = amount of substance discharged (kg/d)
$F_{TOX\,I}$ = toxic weighting factor for substance i

Those Chimiotox units for each substance in an effluent are then combined to arrive at the Chimiotox index (IC) for the effluent, using the following formula. If desired, the same formula could be used to integrate the results for individual effluents, or individual substances, into totals according to types of industry, groups of substances, geographic region, or overall.

$$IC = \sum_{i}^{n} UC_i$$

where: IC = Chimiotox index for a given effluent
UC_i = Chimiotox units for a given substances
n = number of substances

The toxic weighting used for the Chimiotox model makes it possible to do several things on a predictive basis, from physicochemical characterization of the effluent:

- compare the results of characterization of different effluents,
- identify the predominant toxic substances,
- combine the results into a single value for compilation into data bases, and
- obtain an overall picture of the toxic waste situation.

The concept of toxicity coefficients is not new. The procedure derives a relative toxicity coefficient for a pollutant, in a process similar to that for Toxic Units, and similar to that used internationally for polychlorinated dioxins and furans (Section 4.3.2). The Chimiotox units achieve values similar to the Toxicity Emission Factor (Section 4.3.3) in two steps, by first calculating the toxic weighting factor which rates the toxicity of a substance by a unitless factor, and second by incorporating the load discharged.

Toxicity was weighted for long-term effects on human health and aquatic life, for purposes of the St. Lawrence Action Plan. Therefore it was necessary to adopt one or more compendia which assigned quality guidelines to each substance according to the use or environmental compartment to be protected. The source selected was the Cahier des critères de qualité de l'eau, produced by the Ministère de l'environnement du Québec (MENVIC, 1990). Two sets of criteria are normally used by Chimiotox:

(1) contamination of aquatic organisms; and
(2) chronic toxicity to aquatic life.

The first criterion is the aqueous concentration of a substance, to which aquatic organisms can be exposed without bioaccumulating the substance to levels harmful to human health, if humans consumed the aquatic organisms. The second criterion is the usual one for no deleterious effects on aquatic life with indefinite exposure. The more stringent of the two guidelines for a given substance is used to calculate the toxic weighting factor. (In fact, criteria are also considered for acute toxicity to aquatic life, and raw water intended to receive processing for drinking, but they would almost always be less restrictive than the two listed previously.)

An Example of Toxicity Testing as a Predictive Tool in Environmental Planning

The Alberta-Pacific pulp mill is located in northern Alberta, and discharges treated effluent to the Athabasca River. The mill uses mixed hardwood 80% of the time and mixed softwood 20% of the time. In March 1990, following lengthy public hearings, a joint Federal/Provincial environmental review panel concluded that the pulp mill, as originally proposed, should not be built. The major reason was discharge of effluent to the river, and in particular its content of chlorinated organic compounds. Subsequently, Alberta-Pacific proposed a mill that would use chlorine dioxide for 100% of the bleaching, thereby minimizing production of chlorinated organic compounds.

As part of the Environmental Impact Assessment (EIA) for the modified mill, the company sought to demonstrate that the treated effluent would be nontoxic to aquatic life. A predictive assessment was done in the laboratory (Holloran *et al.*, 1992), and is summarized here under its four components.

- Bench-scale pulping and bleaching trials, designed to simulate the processes and their expected effluents.
- Chemical measurements on the bench-scale effluents to characterize the toxic components.
- Toxicity tests with the effluents from the trials.
- Prediction of toxicity based on the chemical components.

Modelling the toxic components added credibility to the various conclusions that stemmed from the assessment.

- The trials indicated that the new process would produce a lower overall amount of organochlorines, and the toxic contributions of the chemical components would be much lower.

- The overall toxicity would be lower than with conventional bleaching processes with chlorine, and the effluent was not expected to cause acute lethality.

- Degradation rates during bio-oxidation were applied to the components thought to cause toxicity, and it was predicted that after secondary treatment, the effluent would not show sublethal toxicity.

The toxicity of the combined components was modelled by "toxic equivalence", assigning a factor to each chemical substance, based on its known toxicity compared to a standard toxic substance. That is the same as the internationally accepted technique used for assigning *toxicity equivalency factors* to the various isomers of chlorinated dioxins and furans (Section 4.3.2), is essentially the same as the "toxic weighting factors" used in Chimiotox (Appendix B), and is a variant of the Toxic Units concept (Section 4.3.1).

Bench-scale bleaching trials and chemical characterization of effluents. The trials were designed to generate effluents from each stage of bleaching, so that it would be possible to measure their toxicity and chemical characteristics. A range of hardwood and softwood feeds were used. The trials yielded estimated loads of toxic compounds per air-dried tonne (Adt) of pulp produced. The combined bleach-plant effluent was mixed to reflect the expected water use and recycling, i.e., component concentrations were converted to equivalent concentrations in untreated total mill effluent using the design mill discharge rate of 46 m^3/Adt. The effect of secondary treatment was estimated on the expected removal efficiencies for phenolics, resin acids and fatty acids. (The efficiencies were based on experience elsewhere. It should be mentioned that when the mill was operating on hardwood, resin acids would not be expected as a component of the waste.)

The effluent was chemically characterized by measuring the predominant individual chlorinated and nonchlorinated phenols, resin acids (if present), and fatty acids, the principal toxic components in many pulp mill effluents. Trials with the proposed process were compared with two trials using conventional cooking and bleaching with hardwood and softwood.

Measured toxicity. Five pilot-plant effluents were tested for sublethal toxicity to *Ceriodaphnia dubia* (EC, 1992a). In those tests, the simulated bleach-plant effluent was diluted to reflect its contribution to the proposed mill effluent.

Measured TOECs for three trial runs with softwood were represented with statistical adequacy by one value, 20.2% concentration of effluent. For two runs with hardwood, toxicities were represented by 40.5% concentration. Since those results are for bleach-plant effluents diluted to concentrations expected in the untreated total mill effluent, therefore the dilutions required in the receiving water to eliminate sublethal toxicity would be 5-fold for softwood and 2.5-fold for hardwood.

Predicting combined toxicity of the components. The major chemical components of the trial effluents were assigned factors for *toxic equivalence*. The factor was obtained by comparing the chemical's known toxicity to that of the most toxic member of its class, i.e., the TEQ method used for dioxins and furans (Section 4.3.2).

For the first category of substances, **chlorinated phenolics**, toxic equivalence factors were assigned on the basis of a report later published by Kovacs *et al.* (1993), with pentachlorophenol as the base. Literature for sublethal effects had been used to assign the toxicity factors, with some assumptions based on QSAR [1]. For example, if a particular substance had 1/4 the toxicity of pentachlorophenol, it would be assigned a toxic equivalence factor of 0.25. Holloran *et al.* (1992) added into this first category (chlorinated phenolics) certain unchlorinated substances (phenol, cresol, guaiacol, catechol, and veratrol) by assigning each substance a factor of 0.001.

For nine **resin acids**, the second category of components, Kovacs *et al.* (1993) assigned factors ranging generally from 0.2 to 0.75, relative to a value of 1.0 for the most toxic one, isopimaric acid. For ten **fatty acids**, the third category, they assigned similar factors to 10 fatty acids, relative to a value of 1.0 for the most toxic, chlorostearic acid. The data base for the resin and fatty acids was not detailed.

Concentrations of the various substances in the effluents of the trial runs were multiplied by their factors, to obtain toxicity equivalents, i.e. "toxicity" concentrations that were unitless. Holloran *et al.* (1992) summed toxicity within each of the three classes of chemicals, but did not add toxicity between classes. As in toxic units, if the sum were 1.0 or greater, one would expect toxicity to be seen. In all test runs the sums for each of the three categories were relatively low in terms of acute lethal toxicity. The highest values were 0.35 for phenolics in one run, and 0.25 for resin acids in another run. Since the three individual categories were well below 1.0, no acute lethality was expected from those compounds.

For sublethal toxicity, the prediction can be compared to that actually measured for the effluents of the five trial runs. Holloran *et al.* (1992) did calculations similar to those for lethality, extrapolated to a sublethal basis, but again did not add together the contributions from the three categories of chemicals.

---

[1] Kovacs *et al.* (1993) took pentachlorophenol as the most toxic base, and estimated the toxicities of phenolics with other degrees of chlorination: other pentachlorinated phenolics 1.0; tetrachloro, 0.48; trichloro, 0.23; dichloro, 0.11; monochloro, 0.05. These toxicity factors showed general agreement with values already published as Canadian water quality guidelines (CCREM, 1987). The Canadian guidelines assigned concentrations (μg/L) according to the degree of chlorination of five categories of phenols, and if recalculated in the same way as done by Kovacs *et al.* (1993), the corresponding five factors would be 1.0, 0.5, 0.028, 0.4, 0.1. There is exact agreement between the factors of Kovacs *et al.* and those of CCREM for tetrachlorophenols, while the other pairs differ by factors of 2, 4, and 8. The general agreement must be considered good, although an examination of the reasons for the larger differences might be in order.

A more useful picture can be obtained by doing that addition and obtaining a single value of total toxic equivalence for each trial run [2]. Doing that, the measured toxicity of the five effluents was, on average, 2.1 times as toxic as predicted by the sum of individual toxicities. The individual values for observed/predicted IC50s were 20/23%, 20/35%, 20/61%, 40/82%, and 40/>100%, i.e., ratios of 1.1, 1.7, 2.0, 2.6, and 3.0 times as toxic as predicted. The correlation is fairly reasonable considering the usual variation in aquatic toxicity and chemical measurements (Section 2.6; Sprague, 1985) and the exploratory nature of the work.

Predicting the effects of effluent. As mentioned previously, one conclusion was that there would not be acute lethal effects, even without secondary treatment of the effluent. Obviously there would be sublethal effects of untreated effluent, judging from the toxicity tests with *Ceriodaphnia*.

The effects of secondary biological treatment on the removal of sublethal toxicity were evaluated through a mass balance based on the expected removal efficiencies of chlorophenolics, resin acids, and fatty acids. The analysis is presented in Figure C-1, which shows the removal efficiency required to reach an effluent concentration corresponding to an IC50 of 100% (i.e., no sublethal toxicity) plotted as a function of the effluent dilution needed to reach the same concentration. Hardwood and softwood effluents are shown separately. Also shown are the expected ranges of removal efficiencies for chlorinated phenolics and for resin and fatty acids. For both effluent types, the expected removal efficiencies are somewhat greater than the efficiencies required to eliminate sublethal toxicity in the effluent. Thus, it was predicted that the mill effluent, after secondary treatment, would not exert sublethal toxicity from chlorophenolics or resin and fatty acids.

The bench-scale exercise, although fairly complex, provides an example of attempting to plan ahead for toxicity control. If successful, such an exercise could help to avoid mistakes in designing an industrial operation at a given site, which might be extremely costly to remedy after the event.

---

[2] The sum of contributions by phenolics, resin acids, and fatty acids was somewhat more complicated than simple addition. An appropriate weighting factor was required for the relative toxicity of the reference substance in each category. A chronic/acute ratio of 0.1 was appropriate (Bonsor *et al.*, 1988; Sections 4.2.1, 4.2.2) instead of 0.2 which had been used by Holloran *et al.* (1992).

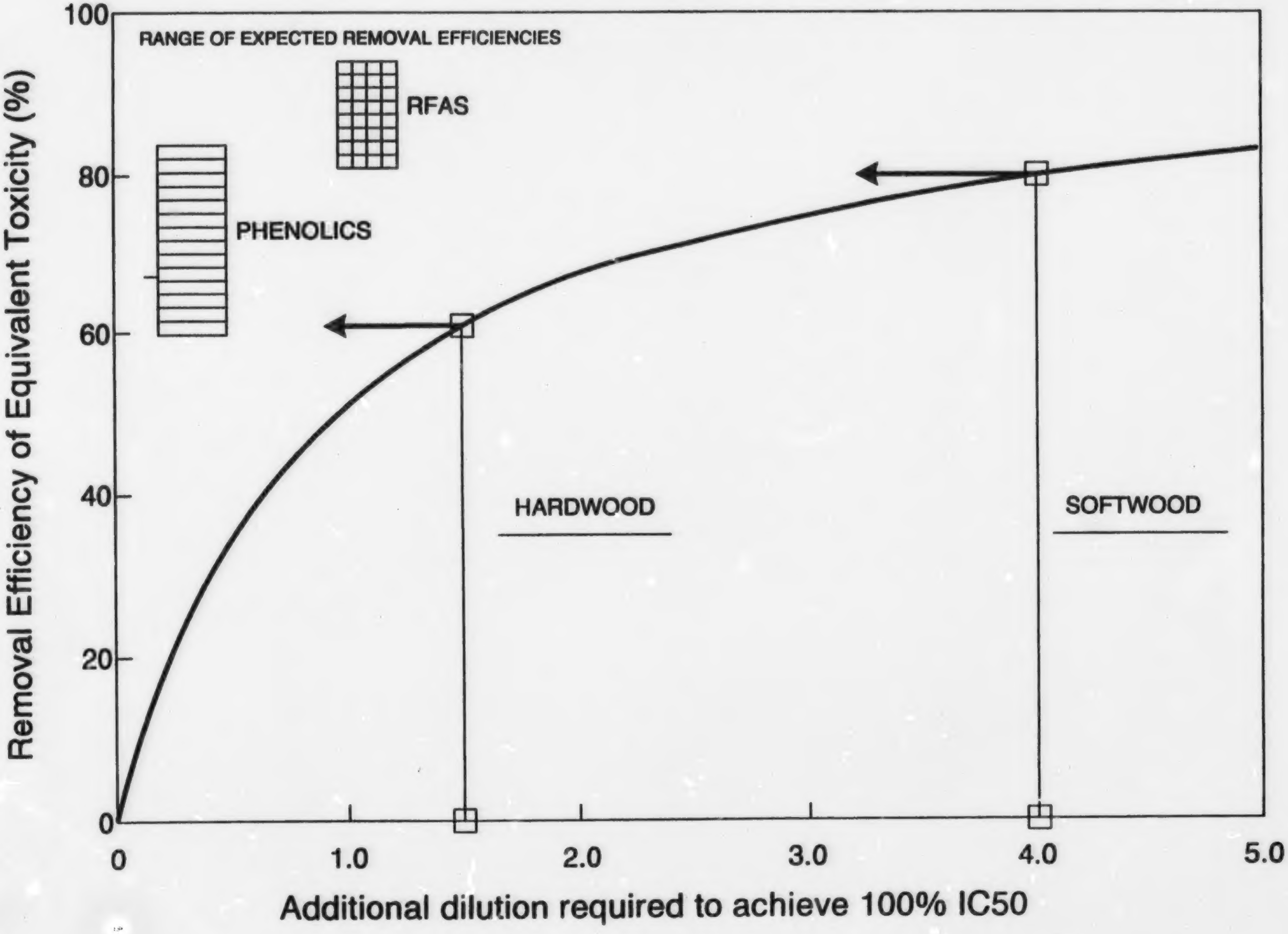

Figure C-1 **Estimated Dilution Required to Achieve no Sublethal Effects, and Efficiency of Removal of Toxic Substances by Secondary Waste treatment.** RFA signifies resin and fatty acids. The horizontal axis represents the *additional* dilution required to achieve no sublethal effect, based on the toxicity tests on effluent from the bench-scale trials. (From Holloran *et al.*, 1992.)

Example Calculations for Predicting Potential Zones of Sublethal Effect in an Effluent Plume

A mill discharges treated effluent to a small river. Plume delineation studies show that at low river flows, the 1% effluent isopleth is approximately 1250 × 250 m and the 10% effluent isopleth is about 50 m wide and extends 250 m downstream from the mid-channel diffuser (upper panel, Figure D-1). A parallel scenario for discharge into a lake is shown in the lower panel of Figure D-1.

Example calculations for toxicity tests on the effluent are shown in Table D-1. There were four estimates of IC25 for each of three species, *Selenastrum, Ceriodaphnia*, and fathead minnow. For samples A and C the IC25 for fathead minnows was reported as >100. Logarithmic means and standard deviations have been calculated for each of the three species based upon the four IC25s. The compensation factor for samples of 4 has been taken from Table D-2. For 95% level of confidence, the compensation factor is 5.369. The calculations as shown below yield the Predicted Minimum Toxic Concentration (PMTC) for each species, which would only be expected to occur in about 5% of toxicity tests. The recommended statistical procedures are appropriate for the numbers of observations dealt with here, which might vary from 4 to 10 data points by the end of the second cycle of EEM.

Table D-1 Example Calculations

Species	IC25, % effluent				geometric mean (g.m.)	logarithm of g.m.	logarithmic standard deviation(s)
	A	B	C	D			
Selenastrum	15	17	25	30	20.9	1.3204	0.1408
Ceriodaphnia	21	92	33	61	44.4	1.6476	0.2837
Fathead minnow	>100	60	>100	72	81.1	1.9089	0.11008

NOTE: If an IC25 is greater than 100%, as with fathead minnows, it is entered into calculations as 100%. If all IC25s are >100, then the standard deviation of the logarithmic values(s) will be zero, and PMTC = 100%.

Calculations for *Selenastrum*. For $n = 4$ measured IC25s, for 95% confidence, the compensation factor *(k)* from Table D-2 is 5.369

The predicted lower limit for 95% of the population of logarithmic IC25s, with 95% confidence, is:

$$\begin{aligned} \text{log PMTC} &= (\text{logarithm of g.m.}) - ks \\ &= 1.3204 - 5.369 * 0.1408 \\ &= 0.5643 \end{aligned}$$

$$\begin{aligned} \text{PMTC} &= 10^{0.5643} \\ &= 3.67\ \% \end{aligned}$$

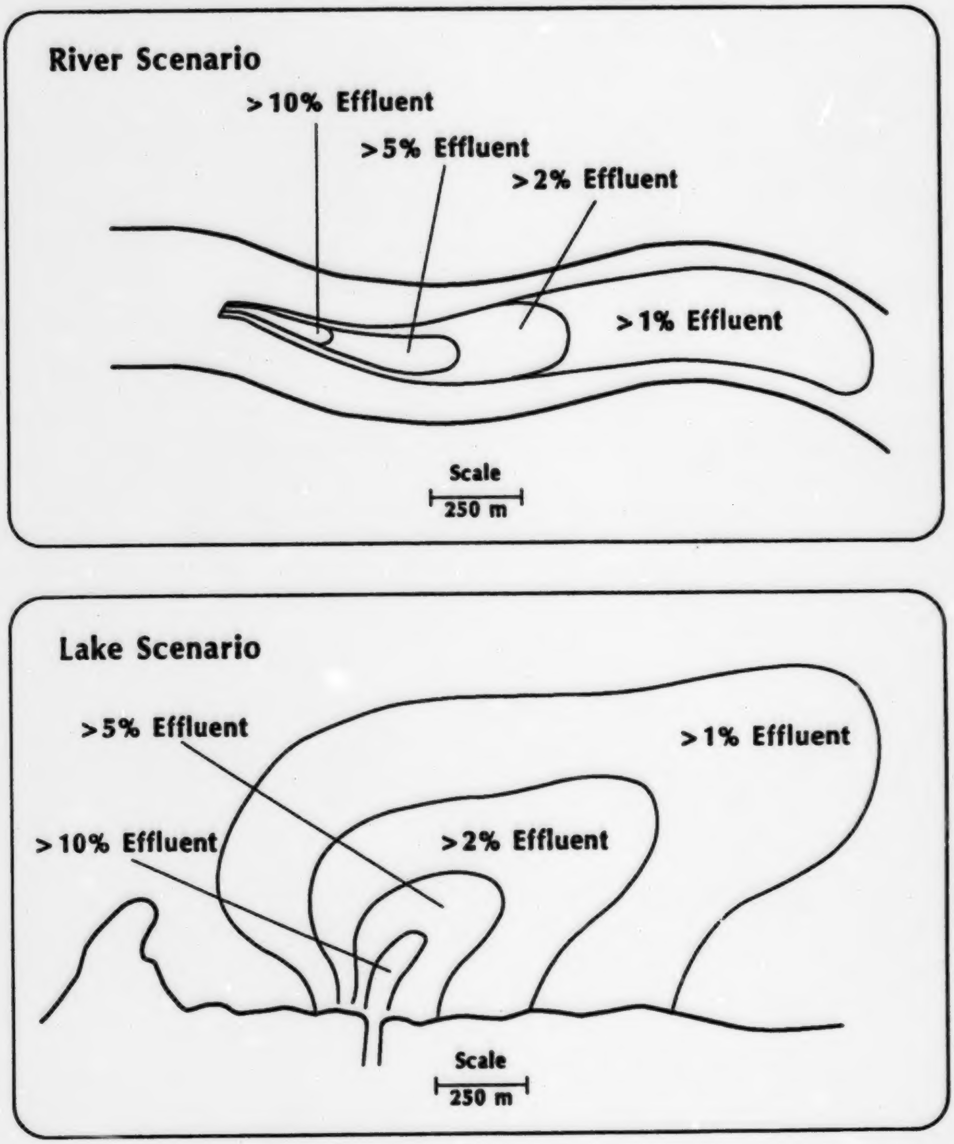
River Scenario
>10% Effluent
>5% Effluent
>2% Effluent
>1% Effluent
Scale
250 m
Lake Scenario
>5% Effluent
>1% Effluent
>10% Effluent
>2% Effluent
Scale
250 m

Figure D-1 Diagrams Illustrating Examples of Delineation of Plumes in a River and a Lake.

In this example, the estimate of PMTC is considerably lower than the lowest of the four measured IC25s for *Selenastrum*, and would be used as a conservative value for evaluating the zone of potential sublethal effect. Use of this statistical technique for estimating PMTC would provide a major incentive for completing at least four sublethal tests, i.e., no failed tests. With only three tests, the compensation factor from Table D-2 becomes much larger, and with only two tests the factor is enormous. If more toxicity tests were done, however, such as eight tests from the first two cycles, the compensation factor would become considerably smaller, thus narrowing the gap between mean measured IC25 and the PMTC. For example, if eight tests with *Selenastrum* had yielded the same logarithmic mean and standard deviations, the PMTC would have been 7.6%, becoming much closer to the lowest value actually found in tests.

Mapping PMTCs. The PMTCs are 1.3% for *Ceriodaphnia*, 3.7% for *Selenastrum*, and 21% for fathead minnow, indicating that the zone for potential sublethal effect in a river or lake would be greatest for *Ceriodaphnia* and smallest for fathead minnow. The lowest PMTC is for *Ceriodaphnia*, as the result of the variable test results and the larger standard deviation.

The size of a potential sublethal effect zone could be shown by locating the PMTC isopleth on a map of the plume area as shown in Figure D-2. The zone of potential maximum effect for *Ceriodaphnia* would be about 1500 m long, and that for *Selenastrum* would be about 720 m long and 150 m wide. The predicted potential effect zone for fathead minnows would be limited to a length of about 100 m and width of about 40 m. Only for fathead minnows would the predicted zone of potential sublethal effects be within the limits defined for mixing zones by Quebec.

Table D-2 **Compensation Factors for Lower Limits of 95% of the Population** (taken from Eisenhart *et al.*, 1947; also available, with short discussion, in: Dixon and Massey, 1969)

n	Confidence Level		n	Confidence Level	
	95%	99%		95%	99%
2	32.019	160.193	15	2.480	2.945
3	8.380	18.930	16	2.437	2.872
4	5.369	9.398	17	2.400	2.808
5	4.275	6.612	18	2.366	2.753
6	3.712	5.337	19	2.337	2.703
7	3.369	4.613	20	2.310	2.659
8	3.136	4.147	25	2.208	2.494
9	2.967	3.822	30	2.140	2.385
10	2.839	3.582	40	2.052	2.247
11	2.737	3.397	50	1.996	2.162
12	2.655	3.250	60	1.958	2.103
13	2.587	3.130	75	1.917	2.042
14	2.529	3.029	100	1.874	1.977

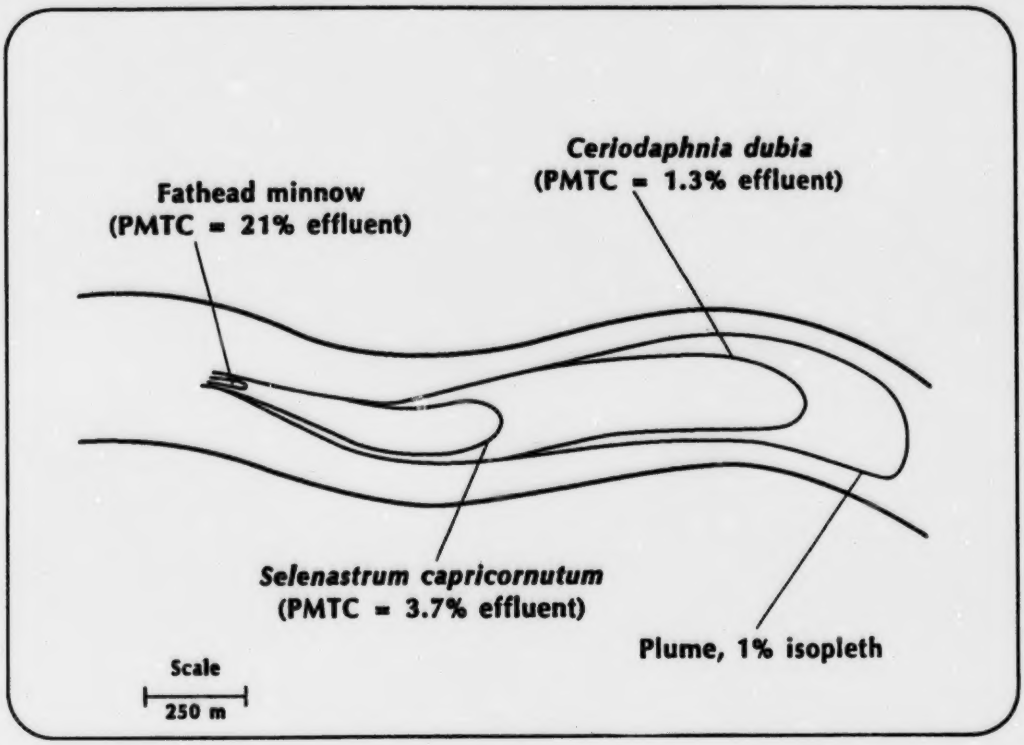
Ceriodaphnia dubia
(PMTC = 1.3% effluent)
Fathead minnow
(PMTC = 21% effluent)
Selenastrum capricornutum
(PMTC = 3.7% effluent)
Plume, 1% isopleth
Scale
250 m

Figure D-2 Isopleths of the Predicted Minimum Toxic Concentrations (PMTC) for Three Species of Test Organisms, Plotted onto the Plume in a River, Taken from the Upper Panel of Figure D-1.